UCLouvain
Economics School of Louvain

Migration, Culture and Politics

Riccardo Turati

7^{th} January 2020

Composition du Jury:
Frédéric Docquier (*UCLouvain*) (Promoteur)
Fabio Mariani (*UCLouvain*)
William Pariente (*UCLouvain*)
Giovanni Peri (*University of California, Davis*)
Hillel Rapoport (*Paris School of Economics*)

Président du Jury:
Fabio Mariani (*UCLouvain*)

"Research is a collective action"

David de la Croix,
PhD Public Defense of Elisa Rizzo,
27^{th} November 2018

Acknowledgements

This thesis is the product of five fantastic years, where I had the opportunity and the grace of meeting fantastic and passionate people who helped me and taught me, or simply sharing their friendship with me, which made those years a unique journey. I really want to thank you all!

I want to start from Frédéric Docquier, my supervisor, who made this path possible. I want to thank him for his enthusiasm related to research, in general, and the migration issue, specifically. Indeed, it was extremely easy to learn, work and do research with someone like him: it made the whole research activity a fascinating world. Moreover, I want to thank him for always encourage me on my research, and also for all the opportunities that he made possible, like spending some of my PhD abroad or sharing our projects with other co-authors. Indeed, my desire to continue on doing research is extremely related with the experience that I have done with him, as student and co-author. Finally, I want to thank Frédéric also for all the discussions and his easygoing approach, which helped me through my PhD experience.

I want to thank Giovanni Peri and Hillel Rapoport, for hosting me as a visitor, and for sharing their time and knowledge with me. Again, working with them and learning from them expanded my views on research and it was an amazing experience. Moreover, I want to help for their guidance and their participation, comments and helps on my whole PhD thesis.

As co-authors and friends, I want to thank Simone Moriconi, Jerome Valette and Dany Bahar. It was really nice to meet you and working with you. I also thank Thomas Baudin, Marion Mercier and Joel Machado for their kind words and support. As part of the jury, I want to thank Fabio and William, for their comments and kind words during the private-defense. Their questions and comments indeed improved the overall quality of my thesis.

I think that one of the several graces of my life was to do my PhD here in Belgium, in particular at IRES. I want to thank all the people that helped to make the IRES building more than a just working-place, but really something familiar and where is possible to invest

and build something. First, I want to thank all the professors that contributes to the life of this place, with their enthusiasm and participation, and from which I had the possibility to learn: Luca, David, Michel, Fabio, Gonzague, William, Bruno, Muriel, Vincent and Joseph. I want to thank all the administrative staff, which helped me in more than one situation: Claudine, Sébastien, Severine, Stephanie, Marie, Solange, Isabelle and Geraldine. A great thank goes to all the IRES-fellas, past and actual for all the shared lunches, PhD aperos, and for their friendship: Adam, Charlotte, Ehui, Douglas, Francesco, Keiti, Guillermo, Charles, Boris, Dalal, Thu Hien, Hendrik, Yannik, Giulio, Pauline, Christoph, Michal, Daniele, Elisa, Annalisa, Fabio, Nathan, Arno, Charles, Leo, Coraline, Esmeralda, Francesca, Esther, Stefanjia, Elsa, Markus, Luigi, Francoise and Zainab. A special thanks goes to François and Juliana for all the dinners, trips and shared moments, and with which I share a tie that is closer to brotherhood, and to Elisabetta, for all the time and moments shared and thanks to whom I am really grateful for all I have.

I also want to thanks all my friends made in Belgium, California and France, which makes this journey even more fantastic: Giorgio, Laura, Zanna, Paolo di Gent, Paolo di Roma, Rascal, Chicco, Grisu, Giulio, Carolina, Francesca, Anna, Laura, Silvia, Cate, Simone, Giulia, Davide, Ana, Paolo, Gila, Peuven, Giggi, Pimer, Angela, Mauro, Cicco, Marta, Sappo, Marghe, Demetrio, Luca, Marie-Sophie, Gerlinho, Federico, Paul, Caterina, Genevieve, Michele, Stefano, Wendy, Uncle Thom, Sue, Tiana, Eric, Scara, Filo, Don Giovanni, Martino, Helene, Alexandre, Cecile, Mathieu, Francesca, Johanna, Anna Zed, Marco, Fede, Paolino, Maite, Francesca, Milena, Ciccio Della, Jack B. and Giacomo. Particularly, I want to thank my actual and old housemates, for their friendship along those years sharing the everyday life: Azzi, Pucci, Michele, Emanuele, Luciano, Tortello, Pippo, Dada, Manau, Axel, Johannes and Luca.

Finally, I want to thank my family, who supported me, even though I was barely at home or present, due to my complete lack of organization. I want to thank my mom Cristina, my father Franco, my brothers and sisters-in-law Pietro, Chiara, Marco and Lucia, and also my nephews Michele, Tommaso, Letiza, Cecilia, Giovanni e Giuseppe, for all the moments shared together. Until now was a fantastic journey, let's see what the future will reveal! Thank you all!

Contents

General Introduction **1**

1 Birthplace Diversity and Economic Growth **5**

1.1 Introduction 6

1.2 Birthplace diversity in the US states (1960-2010) 10

1.2.1 The Birthplace Diversity Index 11

1.2.2 Data and stylized facts 13

1.3 Empirical Strategy 16

1.3.1 Benchmark Specification 16

1.3.2 Identification Strategy 20

1.4 Results 22

1.4.1 Pooled OLS and FE Regressions 22

1.4.2 Robustness checks 25

1.4.3 Dealing with endogeneity 32

1.4.4 Channels of transmission 34

1.5 Conclusions 41

1.6 Appendix 43

1.6.1 Descriptive statistics and data sources 43

1.6.2 Epidemiological effects 66

1.6.3 IV: zero- and first-stage regressions 69

1.6.4 Self-selection of immigrants . . . 74
1.6.5 System GMM regressions . . . 77
1.6.6 Adjusted diversity index . . . 80
1.6.7 Analysis for the OECD countries . . . 82

2 Birthplace Diversity and Economic Complexity 89

2.1 Introduction . . . 89
2.2 Data and Stylized Facts . . . 93
2.2.1 Immigration and Diversity Indexes . . . 93
2.2.2 The Economic Complexity Index . . . 96
2.2.3 Stylized facts . . . 100
2.3 Empirical Strategy . . . 103
2.3.1 Identification Strategy . . . 106
2.4 Main Results . . . 108
2.5 Robustness . . . 115
2.6 Discussion . . . 120
2.7 Conclusion . . . 126
2.8 Appendix . . . 127
2.8.1 Harmonization of DIOC and ADOP data . . . 127
2.8.2 Additional Tables . . . 128

3 Skill of the Immigrants and Vote of the Natives 137

3.1 Introduction . . . 138
3.2 Data and Summary Statistics . . . 143
3.3 Empirical Specification and Identification . . . 156
3.4 Baseline Results . . . 162
3.5 Heterogeneity in Individual and Regional Voting Responses . . . 167

3.5.1 Estimates of Differential Individual Effects 168
3.5.2 Estimates of Differential Regional Effects 170
3.6 Effects on Turnout and on Longitudinal Individual Preferences 172
3.6.1 Decomposing the Effect: Turnout and Voting Preferences 172
3.6.2 Decomposing the Effects: Estimates 175
3.7 Simulations of Nationalist Voting Scenarios 179
3.8 Conclusions 187
3.9 Appendix 189
3.9.1 Additional Tables 189
3.9.2 Descriptive statistics and data sources 189
3.9.3 Principal Component Analysis 199
3.9.4 Parties' names and political preferences 200
3.9.5 Regional level Mechanisms 209
3.9.6 Robustness Checks by regional economic conditions 210
3.9.7 Vote for Party Families 212
3.9.8 Effects on Voting 213
3.9.9 Nationalism and Migration attitudes 214
3.9.10 Individual Panel Analysis 215
3.9.11 Simulations of different scenarios 216

4 Immigration and Voting for Redistribution 220

4.1 Introduction 221
4.2 Data and Summary Statistics 225
4.3 Empirical Specification and Identification 235
4.3.1 Impact on Voters' Preferences 235
4.3.2 Impact on Parties' Platforms 239

4.4 Results on Voters' Preferences 241
4.4.1 Heterogeneity in Individuals' Voting Response 244
4.5 Results on Parties' Political Manifesto 248
4.6 Conclusion 251
4.7 Appendix 253
4.7.1 Additional Tables & Figures 253

5 Emigration and Cultural Selection **259**
5.1 Introduction 260
5.2 From theory to data 263
5.3 Empirical Strategy 271
5.4 Results 275
5.4.1 Logit Regressions 276
5.4.2 Robustness Analysis 279
5.5 Discussion and policy issues 290
5.6 Conclusions 293
5.7 Appendix 294
5.7.1 List of preferred destinations by country of origin 294
5.7.2 Aggregate cultural indices 296
5.7.3 Descriptive Statistics 298
5.7.4 Robustness Analysis 298
5.7.5 Matched Samples 302

6 Connectedness and the diffusion of Culture **305**
6.1 Introduction 306
6.2 Data and Summary Statistics 311
6.2.1 Measuring Connectedness 312

6.2.2 Measuring Cultural Traits 314
6.3 Empirical Strategy 322
6.3.1 Benchmark Specifications 322
6.3.2 Econometric Issues: dealing with selection 325
6.4 Results 329
6.4.1 Benchmark Results 329
6.4.2 Dealing with Selection 336
6.5 Cultural convergence/divergence and mechanisms 344
6.5.1 Cultural Convergence or Divergence 344
6.5.2 Mechanisms and channels of transmissions 351
6.6 How much does it count? 358
6.7 Conclusions 362
6.8 Appendix 364
6.8.1 Matching: Tables and Figures 364
6.8.2 Matching: Sensitivity Analysis 368
6.8.3 Descriptive Statistics 370
6.8.4 Active Religious participation and Discrimination 373
6.8.5 Additional Robustness 376
6.8.6 Controlling for General Connectedness 381
6.8.7 Selection on Unobservables 383
6.8.8 Matching Results 384
6.8.9 Cultural Convergence or Divergence 388
6.8.10 Connectedness Diffusion 392
6.8.11 Individual Heterogeneity 394
6.8.12 Differential effects based on connection location 396
6.8.13 How much does it count? 398

General Introduction **403**

Bibliography **404**

List of Figures

1.1 Trends in birthplace diversity in the US, 1960-2010 13

1.2 Cross-state differences in birthplace diversity, 1960-2010 average index . . 14

1.3 Marginal effect of $Mig\ Div_{s,t}$ on $log(y_{s,t})$ Results for different age-of-entry thresholds (1970-2010) 30

1.4 Augmented Diversity Index 40

1.A1 US immigration rate, 1960-2010 (as percentage of total population) . . . 47

1.A2 Trends in birtplace diversity in the US states 48

1.A3 Diversity among immigrants ($MD^A_{s,t}$) in the US states 49

1.A4 Cross-state correlation between epidemiological terms and GDP 67

1.A5 Difference between predicted and actual diversity indices 73

1.A6 Trends in birthplace diversity in the OECD member states, 1960-2010 . 83

1.A7 Diversity among immigrants ($MD^A_{s,t}$) in the OECD countries 84

1.A8 Marginal effect of diversity conditional to the share of college graduates in the 2000 immigrant stock 86

2.1 Birthplace Migration Diversity 2000 95

2.2 Economic Complexity Index 2000 100

2.3 Evolution of Economic Complexity and Migrants Birthplace Diversity . . 101

2.4 Economic Complexity and Migrants Birthplace Diversity 102

3.1 ESS vs. Actual Data on Voting 149

3.2 Salience and Preference for Nationalism across Parties 150
3.3 Nationalism by Party Family 151
3.4 Nationalism by Left-Right Self-Declared Political Position 152
3.5 Average Nationalism across Regions - Top/Bottom 20 153
3.6 Country-specific Nationalism Over Time 154
3.7 Nationalism: Voters or Parties? 155
3.8 Share of High-skilled and Low-skilled Immigrants Over Time 155
3.9 Variation of Nationalism and Skill-Specific Immigration 156
3.10 Nationalism Response to Immigrant Share and Natives' Individual Characteristics 169
3.11 Political Attitudes of Natives and Immigrants 178
3.12 Baseline Simulation Analysis 182
3.13 Simulations for Different Migration Scenarios 184
3.14 Simulation for Different Education Scenarios 185
A1 Nationalism, Salience and parties 208
A2 Nationalism Distribution 215
A3 Robustness Checks 217
A4 Actual and predicted scenarios 218

4.1 Political Preferences for Net Welfare and Public Education after Electoral Results 229
4.2 Share of votes from ESS and Actual Share of Votes by Country Election in 2007-2016 233
4.3 Parties Families - Average position and Variation 234
4.4 Net Welfare State Expansion and Natives' Individual Characteristics . . 245
4.5 Net Public Education Expansion and Natives' Individual Characteristics . 247
A1 ESS and Actual Voting (Regional Level) 254

5.1 Cross-country heterogeneity in gender-egalitarian views 269
5.2 Cross-country heterogeneity in religiosity 269
5.3 Religiosity: estimated $\hat{\beta}$ and its 95% interval of confidence by country . . 285
5.4 Gender-egalitarian views: estimated $\hat{\beta}$ and its 95% interval of confidence by country) 286
A1 Principal Component Analysis - Loadings plot 296
A2 Factor Analysis - Loadings plot 297

6.1 Geographical distribution of Connected People 313
6.2 Distribution of Cultural Traits 316
6.3 Matching Results - Average effect of Connectedness 342
6.4 Within Regions Cultural Heterogeneity - Overall Heterogeneity 349
6.5 Within Regions Cultural Heterogeneity - Between Groups Heterogeneity . 350
6.6 Relative deviation from the No-connection Scenario - Country Level . . . 360
A1 Probability Score Matching Distribution. 365
A2 Matching Results - Robustness to fake treatment on control group . . . 367
A3 Social Activities and Cultural Traits 382
A4 Average Effect of Connection - Different Matching Algorithms 384
A5 Distribution Covariates after Matching - Social Behavior. 385
A6 Distribution Covariates after Matching - Religiosity 386
A7 Distribution Covariates after Matching - Gender-Egalitarian 387
A8 Relative deviation from the No-connection Scenario - Individual Level . . 398
A9 Average Trait by country of birth - Population and Migrants abroad. . . 402

List of Tables

1.1 Pooled OLS and FE regressions Results by skill group (Dep= $log(y_{s,t})$) . 23

1.2 Robustness of FE regressions for high-skilled diversity (Dep= $log(y_{s,t})$) . 26

1.3 $Mig\ Div_{s,t}$ v.s diversity among "native immigrants" $Nat\ Div_{s,t}$ Results by skill group (Dep= $log(y_{s,t})$) 33

1.4 2SLS regressions under different IV strategies. (Dep= $log(y_{s,t})$) 35

1.5 Transmissions channels for high-skilled diversity IV gravity-like strategy *a la* Feyrer (2019) (Dep= $log(y_{s,t})$) 37

1.A1 List of origin countries (195). 44

1.A2 Sources and definitions 45

1.A3 List of US states and descriptive statistics 46

1.A4 Pearson correlations between diversity measures 50

1.A5 Summary statistics 1960-2010 51

1.A6 Robustness of FE regressions for low-skilled diversity (Dep= $log(y_{s,t})$) . . 53

1.A7 Robustness of FE regressions for high-skilled diversity Alternative subsamples (Dep= $log(y_{s,t})$) 54

1.A8 Robustness of FE regressions for low-skilled diversity Alternative subsamples (Dep= $log(y_{s,t})$) 55

1.A9 Robustness of FE and IV regressions to spatial scale Results by skill group at the CZ level (Dep= $log(Wage^{K}_{CZ,t})$) 56

1.A10 Robustness of FE regressions for high-skilled diversity Ten largest US immigrants group in 2010 (Dep= $log(y_{s,t})$) 57

1.A11 Robustness of FE regressions for low-skilled diversity Ten largest US immigrants group in 2010 (Dep= $log(y_{s,t})$) 58

1.A12 Robustness of FE estimates to alternative specifications and measures Results by skill group (Dep= $log(y_{s,t})$) 59

1.A13 Robustness of FE regressions to alternative educational levels (Dep= $log(y_{s,t})$) 60

1.A14 Robustness of FE regressions Results by legal status and skill group (Dep= $log(y_{s,t})$) 61

1.A15 Robustness of Pooled OLS, FE and IV regressions without controls Results by skill group (Dep= $log(y_{s,t})$) 62

1.A16 Robustness of FE regressions to age of entry, 1970-2010 63

1.A17 Regressions with region-by-year fixed effects 64

1.A18 FE regressions with aggregated group of migrants. Results by skill group (Dep= $log(y_{x,t})$) 65

1.A19 Robustness of FE estimates to alternative definitions of the epidemiological term Results by skill group (Dep= $log(y_{s,t})$) 68

1.A20 Zero-stage estimates (PPML): gravity model *a la* Feyrer (2019) 70

1.A21 First-Stage regressions High-skilled (Alternative IV Strategies) 71

1.A22 First-Stage regressions Low-skilled (Alternative IV Strategies) 72

1.A23 Accounting for selection on observables and unobservables (Dep=$log(y_{s,t})$) 75

1.A24 System GMM: Internal instruments Results by skill group (Dep= $log(y_{s,t})$) 78

1.A25 System GMM: External instruments Results by skill group (Dep= $log(y_{s,t})$) 79

1.A26 Robustness to the measure of diversity Accounting for genetic distance between countries (Dep= $log(y_{s,t})$) 81

1.A27 Descriptives statistics for the OECD countries (34) 85

1.A28 Birthplace diversity in a cross-country setting. Regressions for the OECD member states (Dep= $log(y_{s,t})$) 87

1.A29 Zero-stage estimates (PPML): gravity model *a la* Feyrer (2019) 88

2.1 Replication of Alesina et al. (2016) controlling for Economic Complexity . 103

2.2 OLS regression on Economic Complexity Index. 109

2.3 OLS regression on Economic Complexity Index - ECI Terciles 111

2.4 OLS regression on Economic Complexity Index Origin Specific Effect . . 113

2.5 OLS regression on Economic Complexity Index Robustness by previous level of diversity 115

2.6 OLS regression on Economic Complexity Index - Country fixed effects . 117

2.7 Panel Analysis on Economic Complexity Index 118

2.8 IV regression on ECI 119

2.9 Correlations of diversity, diversification, ubiquity, income and complexity. 123

2.10 OLS regression on Diversification and Ubiquity Measures 125

A1 Correlates across different measures of Diversity 128

A2 OLS regression on Economic Complexity Index - Different Diversity Indexes. 129

A3 PPML regression on bilateral migration stocks like AHR. 130

A4 Descriptive Statistics Migration 131

A5 Correlations across birthplace diversity indices 132

A6 Correlations across shares of migrants. 132

A7 Alternative IV Estimates - Shift-share based on 1960 distribution 133

A8 Correlations of diversity and alternative diversification measures 134

A9 OLS regression on Countries Top Products Ubiquity 135

A10 OLS regression on Diversification and Ubiquity Measures 136

3.1 Elections and ESS Rounds by Country and Year 146

3.2 Summary Statistics 148

3.3 Correlation between 2000-2007 Growth of Economic Indicators and Post-2006 Predicted Migrants Growth . . . 161
3.4 Nationalism Intensity and Immigrant Share . . . 163
3.5 Nationalism and Immigrant Share - Including Controls . . . 165
3.6 Voting for Radical Right - Liberal Parties and Immigrant Share . . . 167
3.7 Nationalism and Immigrant Share - Regional Subsamples . . . 171
3.8 Voting Turnout and Immigrant Share . . . 176
3.9 Nationalism and Immigrant Share - Individual Panel Analysis . . . 177
3.10 Immigration and Attitudes towards Politics . . . 179
3.11 Simulated Effect of Immigration/Education Changes on Nationalism: Country-level and EU averages . . . 181
A1 Nationalism Intensity and Immigrant Share, Initial and Final Party Positions . . . 189
A2 Salience of Nationalism and Immigrant Share . . . 190
A3 Nationalism and Individual Characteristics . . . 191
A4 Attitudes towards Migrants and Immigrant Shares - All Native Respondents. . . . 192
A5 Data Sources and definitions. . . . 194
A6 Summary Statistics - Party closeness data . . . 197
A7 Party closeness and party voted: differences (%) . . . 198
A8 ESS data and actual voting (Δ) . . . 198
A9 PCA on favorable political position . . . 199
A10 PCA eigenvectors . . . 199
A11 PCA on measures of salience . . . 199
A12 PCA eigenvectors . . . 199
A13 Partys' names and political preferences . . . 200
A14 Nationalism - Regional characteristics - Migrants Origin . . . 209

A15 Nationalism - Subsample analysis based on unemployment rate quartiles . 210

A16 Nationalism - Subsample analysis based on Unemployment and Unemployment Variation . . . 211

A17 Voting Radical Right and Liberal parties - Voters Education . . . 212

A18 Voting - Voters political orientation . . . 213

A19 Party Closeness - Nationalism and attitudes towards migration . . . 214

A20 Simulations: exercises and robustness checks . . . 216

A21 Actual and predicted Nationalism . . . 219

4.1 Elections and ESS rounds by country and year . . . 227

4.2 Descriptive Statistics . . . 230

4.3 Validation ESS data versus EULFS data . . . 232

4.4 Political Preferences over 2007-2016 - Individual Voters . . . 242

4.5 Political Preferences - Parties . . . 249

4.6 Parties Political Preferences - Robustness . . . 250

A1 Correlation between 2000-2007 Growth of Economic Indicators and Post-2006 Predicted Migrants' Growth . . . 253

A2 Political Preferences Over 2007-2016 - Individual Voters Average Preferences Over 2007-2016 . . . 255

A3 Political Preferences over 2007-2016 - Individual Voters Left to Right Index . . . 256

A4 Net Welfare State - Individual Heterogeneity Analysis . . . 257

A5 Net Public Education - Individual Heterogeneity Analysis . . . 258

5.1 Cultural traits: mean levels by country . . . 268

5.2 Correlation between aggregate cultural GWP, WVS and macro indices . . . 270

5.3 Benchmark regressions . . . 277

5.4 Results for intentions and plans to migrate . . . 278

5.5 Robustness analysis by preferred destination countries 279
5.6 Gender views and migration aspirations towards OECD countries Subsample Analysis 280
5.7 Religiosity and migration aspirations towards OECD countries Subsample Analysis 281
5.8 Analysis on the Matched Sample 283
5.9 Adding interactions with macroeconomic variables 287
5.10 Cultural selection before and after the Arab Spring 289
A1 Top-5 preferred OECD destinations. 294
A2 Top-5 preferred non-OECD destinations 295
A3 Principal component eigenvectors 296
A4 Factor loadings and unique variance 297
A5 Descriptive statistics & Robustness Analysis 298
A6 Pearson correlation between aggregate indices and GWP responses . . . 299
A7 Pearson correlation across variables. 299
A8 Pearson correlation between aggregate indices (religiosity) 299
A9 Pearson correlation between aggregate indices (gender-egal. views) . . . 299
A10 Benchmark Regressions with Gulf Countries 300
A11 Logit regressions - Robustness to different indicators 2007-2011) Dependent = Migration aspirations to OECD/non OECD destinations. . 301
A12 Benchmark Regressions with selected fixed effects. 302
A13 Mahalanobis matching procedure 2007-2011 Matched and non-matched samples 303
A14 Before- and After-Arab Spring Covariate distributions Matched and non-matched samples. 304
6.1 Cultural traits Correlations - GWP and WVS 320
6.2 Cultural traits and Economic preferences Global Preferences Survey . . . 321

6.3 Benchmark Results . . . 330
6.4 Benchmark Results - Robustness . . . 333
6.5 Broad Connections or Household Connections? . . . 334
6.6 Controlling for General Connectedness . . . 337
6.7 Selection on Unobservables . . . 339
6.8 Mahlanobis Matched Sample Analysis . . . 344
6.9 Cultural Convergence or Divergence - Regional subsamples by regional culture . . . 346
6.10 Connectedness Diffusion - Regional subsamples by connected population . 352
6.11 Individuals heterogeneity - Sociodemographic groups . . . 354
6.12 Connectedness and Economic Outcomes . . . 355
6.13 Benchmark Results among Connected People - Connection Location . . . 357
A1 Propensity Score Matching Estimation . . . 364
A2 Sample Means and Standardized Bias - Before and After Matching . . . 366
A3 Sensitivity Analysis - Rosenbaum Bounds . . . 369
A4 List of origin countries (157) . . . 370
A5 Descriptive Statistics . . . 371
A6 Pearson's Correlation across cultural questions . . . 372
A7 Religiosity: Practicing and Religious groups . . . 373
A8 Social Behavior and Gender-egalitarian views: perceived discrimination . 375
A9 Connectedness or Openness - Individual controls . . . 376
A10 Connectedness or Openness - Country subsamples by development and trade exposure . . . 377
A11 Benchmark Estimates - Probit and Ordered Probt Model . . . 378
A12 Analysis on Subcomponents of Social Behavior and Gender-Egalitarian attitudes . . . 379
A13 Benchmark Analysis - Countries Land . . . 379

A14 Connectedness and Culture - Multiple Connections 380

A15 Controlling for General Connectedness - Subsamples 381

A16 Selection on Unobservables - No interaction 383

A17 Cultural Convergence or Divergence - Regional subsamples by regional culture (Median). 388

A18 Cultural Convergence or Divergence - Regional subsamples by regional culture (Quartiles) 389

A19 Cultural Convergence or Divergence - Broad Continents 390

A20 CF by Trait and F_{st} by Cleavage. 391

A21 Country of origin heterogeneity - Network presence (Median). 392

A22 Country of origin heterogeneity - Network presence (Quartiles) 393

A23 Individuals heterogeneity - Natives and Migrants 394

A24 Individuals heterogeneity - Education, same sample of regions 395

A25 Cultural Convergence or Divergence - Country subsamples by connection culture (Median). 396

A26 Cultural Convergence or Divergence - Country subsamples by connection culture (Terciles). 397

A27 Simulations: country values 398

General Introduction

The demographic rise of developing countries, the reduction of transportation and information costs of migration and the relevant economic gap between developed and developing countries make international migration one of the key and rising issues of the next decades. During the last decades, the population of international migrants has considerably increased at the global level. The size of international migrant population reached 258 millions in 2017, while it was only 77 millions in 1960 (United Nations, 2017). Although it is still a moderate amount compared to the worldwide population (3.4% in 2017, United Nations, 2017), the increasing relevance of such topic from political debates to the everyday talks asks for a proper understanding of the effects of immigration on receiving countries and on the implication and the causes of emigration for sending countries. The aim of this thesis is to understand through empirical methods not only the role of migration on the economic development of different countries, but also its role in affecting voting behaviours and cultural beliefs and traits. Since migrants are not just workers, my research aims at providing novel evidence of the impact of international migration outside the well-known and relevant labour markets effects. My thesis evolves on three main issues. The first part investigates the role and the economic implications of a more diverse society, due to the rise of immigration; the second part explores the impact of immigration on Europeans voters' preferences; the third and last part contributes to a better understanding of the relationship between immigration, emigration and individuals cultural traits and beliefs.

First, my thesis sheds some new light on the economic implication of a more diverse society. Coming from distant societies, immigrants bring with them their culture, skills and norms which increase the diversity within a society. However, from an economic point of view, more diversity can bring positive and negative effects. On one hand, diversity in languages and traditions could make communication and coordination between individuals more difficult and costlier. On the other hand, diversity is associated with an increase in variety of competences and skills, which can conduct to new products, activities and innovation. In the first chapter of my thesis, co-authored with Frédéric Docquier, Jerome Valette and Chrysovalan-

tis Vasilakis, I investigate the impact of migration diversity (defined as birthplace diversity, which measures the probability to draw from the immigrants population two individuals born in two different countries) on the economic growth of the 51 US states from the 1960 to 2010. We show that diversity among college-educated migrants positively affect states economic growth. We provide evidence of skill complementarities between workers trained in different countries, particularly in better labour market outcomes for native workers and in higher productivity in the R&D sector. Moreover, the gains from diversity are maximized when immigrants originate from economically or culturally distant countries (but not both), and when they acquired part of their secondary education abroad and their college education in the US. In the second chapter of my thesis, co-authored with Dany Bahar and Hillel Rapoport, we investigate even further the effect and role of diversity on countries' economic development, by studying the relation between countries' population diversity and economic complexity. Borrowing the definition and measure of economic complexity by Hidalgo and Hausmann, 2009, which is defined as country embodied knowledge and productive know-how, we show in a cross-country analysis over 100 countries on the 1990-2000 period that diversity among migrants is positively related with countries economic complexity. Using alternative indicators of diversity and decomposing our measure of economic complexity, we show that the effects are enhanced on developing countries and when the number of origin countries represented in the immigrants population increases (i.e. new available knowledge and productive know-how). Finally, we provide evidence suggesting that birthplace diversity boosts economic complexity by increasing the diversification of the host country's export basket.

Second, my thesis contributes to a better understanding of the effect of immigration on natives voting preferences. The fact that immigration is a salient political topic for voters in developed countries is clearly depicted by the rise of populist/nationalistic parties in several European national elections and also by the Brexit decision of British citizens. However, immigrants are not just a homogeneous group of individuals: they have different characteristics, competences and also a degree of relativeness with natives in the destination countries. Are natives voting behaviors differently affected by the type of immigrants they receive in their region? Investigating this issue with Simone Moriconi and Giovanni Peri over 12 European countries over the 2007-2016 period and using individual level data on individual voting preferences and measures of parties' political platforms, we answer to such question affirmatively: the effect of immigration on natives voting preferences is education-specific. In the third and fourth chapters of my thesis we show that the presence of highly educated immigrants, who are easily integrated into the labor market and perceived as contributors to the welfare system, have pushed European voters towards less nationalistic and more

redistributive parties. Lowly educated immigration has instead driven voters towards more nationalistic and less redistributive parties. Moreover, we show that the presence of immigrants only influences lowly educated natives voting behavior, while has almost no effect on college educated natives. As expected, the impact of immigration on voting preferences depends both on natives and immigrants education. In the third chapter we show that the aggregate results derive in part from an individual shift towards nationalism associated with less-skilled immigration, greater participation of young voters, and pro-Europe attitudes associated with high-skilled immigration. In the fourth chapter we explore parties reaction to immigration, and we show that inflows of less educated immigrants induce European parties to endorse platforms less favorable to social welfare.

Third, my thesis aims to increase the understanding on the relationship between migration and culture. As the recent increase of international migration population around the world is driven by individuals moving from developing to developed societies, international migrants and natives in the destination countries do not usually share the same cultural values and background. Interacting with new cultures, traditions and individuals can influence cultural aspects both at origin and destination countries. Moreover, cultural features can influence both emigrants' preferences and natives' attitudes towards newcomers, and also individual economic and non-economic preferences and behaviors. In the fifth chapter of my thesis, coauthored with Frédéric Docquier and Aysit Tansel, we show that cultural aspects influence individuals' emigration decisions. Focusing on 17 Middle East and North African countries, we show that emigrants are culturally selected: less religious and more gender-egalitarian individuals are more prone to move towards OECD destination countries. In the sixth chapter of my thesis, using individual level data over 148 countries, I show that having a reliable connection abroad (i.e., connectedness) makes natives culturally different compared to neighbors living in the same region. Sharing the knowledge and experience of people around the world makes individuals more active in the society, gender-egalitarian and religious. I provide also suggesting evidence that connectedness favors cultural convergence across regions, while increases cultural heterogeneity within regions. Exploring the mechanisms, I provide evidence that the effects are precisely estimated among less educated natives and connectedness affects economic outcomes through remittances. Differential cultural effects based on connection's residence are estimated, suggesting a destination-specific transfer of norms. Finally, I show that the effects are sizeable at the global level on social behavior, once simulations based on estimated coefficients are performed. Although robust, gender-egalitarian and pro-religiosity effects of connectedness are limited, albeit not negligible. By moving from their origin countries and interacting with their peers at home, migrants influence cultural traits both in the origin and destination countries.

Overall, my thesis attempts to illuminate and provide novel evidence on the relation of international migration, population diversity, political preferences and culture. Through an empirical approach and investigating potential mechanisms, it contributes to the debate on the actual effects of migration, in the context of increasing human mobility.

Chapter 1

Birthplace Diversity and Economic Growth: Evidence from the US States in the Post-World War II Period

Abstract[1]

This paper empirically investigates the impact of birthplace diversity on economic growth. We use panel data on US states over the 1960-2010 period. This rich data set allows us to better deal with endogeneity issues and to conduct a large set of robustness checks. Our results suggest that diversity among college-educated immigrants positively affects economic growth. We provide converging evidence pointing at the existence of skill complementarities between workers trained in different countries. These synergies result in better labor market outcomes for native workers and in higher productivity in the R&D sector. The gains from diversity are maximized when immigrants originate from economically or culturally distant countries (but not both), and when they acquired part of their secondary education abroad and their college education in the US. Overall, a 10% increase in high-skilled diversity raises GDP per capita by about 6%. On the contrary, low-skilled diversity has insignificant effects.

[1]The paper, coauthored with Frédéric Docquier, Jérôme Valette and Chrysovalantis Vasilakis, has been accepted for publication in Journal of Economic Geography.

1.1 Introduction

Patterns of international migration to industrialized countries have drastically changed since World War II (WW2). On average, the share of foreigners in the population of high-income countries increased from 4.9 to 11.7% between 1960 and 2010 (Özden et al., 2011). This phenomenon has similarly affected the United States (from 5.4 to 13.6%), the members of the European Union (from 3.9 to 12.2%), Canada and Australia (from 15 to 22%). In addition, this change has been predominantly driven by immigration from developing countries; the share of South-North immigrants in the population of high-income countries increased from 2.0 to 8.7% in half a century.[2] The literature investigating the economic impact of immigration on the United States and on other host countries has grown rapidly in the past decades (Borjas, 1994). It sheds light on how the welfare of natives and economic outcomes are affected by the number and characteristics of immigrants. Among the multiple transmission channels, the labor market, fiscal, and market size effects of migration have been abundantly studied, and a growing consensus on how to formalize and quantify them has gradually emerged.[3] Nevertheless, the growing inflow of people coming from geographically, economically and culturally distant countries raises specific issues, as it has conceivably brought different skills and abilities, but also different social values and norms, or different ways of thinking. The macroeconomic effects of birthplace diversity, as well as the channels through which they materialize, are still uncertain.

This paper empirically investigates the impact of birthplace diversity on the macroeconomic performance of US states (proxied by their level of GDP per capita) since WW2. Our analysis combines three innovative features. *First*, we rely on panel data available for a large number of regions over a long period. Our sample covers all US states over the 1960-2010 period in ten-year intervals. This rich data set allows us to conduct a large number of robustness checks, and to better deal with unobserved heterogeneity and other endogeneity issues. This is crucial because economic prosperity and the degree of diversification in production are likely to attract people from different cultural origins (see Alesina and La Ferrara, 2005), implying that causation is hard to establish in a cross-sectional setting. To control for unobserved heterogeneity and reverse causation biases, our paper uses a great variety of geographic and time fixed effects, and combines various instrumentation strate-

[2]Immigration from developing countries accounts for 98% of the 1960-2010 rise in immigration to high-income countries, for 80% in the European Union, for 120% in the United States, and for 150% in Australia and Canada. Trends in immigration to the US are presented in Figure 1.A1 in the Appendix.

[3]See among others, Borjas (2003), Card (2009), Ottaviano and Peri (2012), Storesletten (2000), Dustmann and Frattini (2014), Iranzo and Peri (2010), di Giovanni et al. (2015), Aubry et al. (2016), Burzyński et al. (2018).

gies that have been used in the existing literature. *Second*, we systematically investigate whether the economic effect of birthplace diversity is heterogeneous across skill groups. The costs and benefits from diversity are likely to vary with the levels of task complexity and interaction between workers; meanwhile, high-skilled and low-skilled immigrants are likely to heterogeneously propagate social values and norms across borders. We account for this by using skill-specific measures of birthplace diversity. In addition, taking advantage of the availability of microdata, we compute our indices of diversity for different groups of immigrants (by age of entry or by legal status). *Third*, we conduct an exploratory analysis of the channels of transmission through which birthplace diversity affects economic growth, and of the conditions under which the growth-enhancing effect of diversity is maximized. To do so, we test whether diversity affects the alternative outcomes of interest or intermediate variables influencing the level and/or growth rate of GDP per capita. We also use augmented diversity indices to investigate whether the gains from diversity are influenced by the cultural and economic distances between the origin countries and the US.

Our paper belongs to a recent and growing strand of literature, which considers (i) that culture can be a feature differentiating individuals in terms of their attributes, (ii) that such a differentiation may have positive or negative effects on people's productivity, and (iii) that culture is affected by the country of birth (which determines the language and social norms individuals were exposed to in their youth, the education system, etc.). On the one hand, homogeneous people are more likely to get along well, which implies that cultural diversity may reduce trust or increase communication, cooperation and coordination costs. Moreover, diversity can also be the source of epidemiological effects if immigrants convey the cultural, social and institutional norms prevailing in their origin countries, as argued by Collier (2013) and Borjas (2015). On the other hand, cultural diversity also enhances complementarities across diverse productive traits, stimulating innovations and the collective ability to solve problems since a more diverse group is likely to bring varied solutions to the same problem. Evidence of such costs and benefits has been found in micro studies. For example, Parrotta et al. (2014) investigate the effect of different forms of diversity (by education, age group, and nationality) on the productivity of Danish firms, using a matched employer-employee database. They find a negative effect of worker diversity by nationality on productivity. On the contrary, Ozgen et al. (2014) find that birthplace diversity increases the likelihood of innovations using Dutch firm-level survey data, and Boeheim et al. (2012) find a positive effect of diversity on productivity using Austrian data. In the same vein, Kahane et al. (2013) find a positive effect of diversity on hockey team performance using data from the NHL (the North American National Hockey League).

Contrary to the firm-level approach, the analyses conducted at the macro level account for interdependencies between firms, industries, and/or regions. Existing studies have identified significant and positive effects of cultural diversity on comparative development and on disparities in economic performance across modern societies.[4] Ottaviano and Peri (2006) use US data by metropolitan area over the 1970-1990 period. In their (log of) wage regressions, the coefficient of diversity varies between 0.7 and 1.5. Ager and Brückner (2013) use US data by county during the 1870-1920 period: the coefficient of diversity in the output per capita regressions varies between 0.9 and 2.0. In these two studies, endogeneity issues are solved by using a shift-share method, i.e. computing the diversity index on the basis of predicted immigrant stocks. More precisely, the change in immigration to a region is predicted as the product of the global change in immigration to the US by the regional share in total immigration in the initial year. A more recent study accounting for the education level of immigrants is that of Alesina et al. (2016); it is the most similar to ours. They use cross-sectional data on immigration stocks by education level for a large set of countries in the year 2000, and develop a pseudo-gravity first-stage model to predict migration stocks and birthplace diversity indices. They also identify a positive effect of birthplace diversity in countries with a GDP per capita above the median, and a stronger effect for diversity among college-educated workers. The effect of diversity on the log of GDP per capita is around 0.1 when computed on low-skilled workers, while the effect of diversity among the highly skilled varies between 0.2 and 0.3. Similarly, Suedekum et al. (2014) use annual German data by region from 1995 to 2006. Over this short period, they find a weaker effect of diversity on the log of German wages (about 0.1 for diversity among high-skilled foreigners, and 0.04 for diversity among the low skilled) when fixed effects and IV methods are used.

Our empirical analysis relies on high-quality US census data by state over the 1960-2010 period. The choice of this period is guided by the 1965 amendments to the Immigration and Nationality Act, which led to an upward surge in U.S. immigration and diversity (as in Ottaviano and Peri (2006)). Birthplace diversity is almost perfectly correlated with the state-wide proportion of immigrants, which has increased threefold since 1960 in all skill groups. It is thus statistically impossible to disentangle the effects of birthplace diversity from those of the size of immigration. For this reason, we opt for a benchmark model that includes the immigration rate and a birthplace diversity index pertaining to the immigrant population. In line with Alesina et al. (2016) and Suedekum et al. (2014), we find that diversity among college-educated immigrants is positively associated with the level of GDP

[4] Ashraf and Galor (2013) use the concept of genetic diversity (based on the initial concept of expected heterozygosity, that is the probability of selecting two random people from the population who are genetically different) and find a strong hump-shaped relation with the level of development around the years 1500 and 2000.

per capita; however, diversity among less educated immigrants has insignificant (or weakly significant and much smaller) effects. Another remarkable result is that the estimated coefficient is divided by four when geographic and year fixed effects are included. Still, our estimates are greater than those reported in Alesina et al. (2016) and Suedekum et al. (2014): a 10% increase in high-skilled diversity raises GDP per capita by about 6%. There is converging evidence pointing at the existence of skill complementarities between workers trained in different countries. These synergies result in better labor market outcomes for natives, and in higher productivity in the R&D sector. Our results are robust to the exclusion of some census years, to the set of US states included in the sample, to the measurement of diversity, and to the definition of a high-skilled immigrant. The results hold true when we eliminate states with the greatest or smallest levels of immigration share, states located on the Mexican border, and states with the lowest proportions of immigrants. They are also valid when we exclude undocumented immigrants and those who arrived in the US at a young age. In addition, we find no evidence of an inverted-U shaped relationship as found by Ashraf and Galor (2013) for genetic diversity, or of a negative epidemiological effect *a la* Collier (2013) and Borjas (2015). The growth-enhancing effect of diversity is maximized when immigrants originate from economically or culturally distant countries (but not both), and when immigrants acquired primary and part of their secondary education abroad and their college education in the US.

To address endogeneity issues, we combine placebo tests with IV regressions. As far as the latter are concerned, we consider two instrumentation strategies that have been used in the related literature. The first one is a shift-share strategy *a la* Ottaviano and Peri (2006) which includes the predicted diversity indices based on total US immigration stocks by country of origin, and the bilateral state shares observed in 1960. The second strategy consists in instrumenting diversity indices, using the immigration predictions of a pseudo-gravity regression that include interactions between year dummies and the geographic distance between each country of origin and each state of destination (in line with Feyrer (2019) or Alesina et al. (2016)). In both cases, diversity among college-educated migrants remains highly significant, while diversity among the less educated is insignificant or weakly significant. In the preferred specifications, the coefficient of high-skilled diversity remains around 0.6. Since the average diversity index among college-educated immigrants equaled 0.937 in 2010; hence, increasing diversity from zero to 0.937 increases GDP per capita by 58%. However, in 2010, the high-skilled diversity indices of the US states ranged from 0.797 to 0.976. This implies that, if all US states had the same level of diversity as the District of Columbia (0.976), the average GDP per capita of the US would be 2.33% larger, the coefficient of variation across states would be 2.37% smaller, and the Theil index would decrease by 3.45%. By comparison,

if all US states had the same average level of human capital as the District of Columbia, the average GDP per capita of the US would be 8.32% larger, the coefficient of variation across states would be 9.77% smaller, and the Theil index would decrease by 16.06%.[5] Since the US-state average level of diversity among college-educated immigrants increased by 7 percentage points between 1960 and 2010; this explains a 3.5% increase in macroeconomic performance (i.e. one fiftieth of the total change in the US level of GDP per capita).

The remainder of the paper is organized as follows. Section 1.2 describes our main diversity measures and documents the global trends in cultural diversity since WW2. Section 1.3 describes our empirical strategy. The results are discussed in Section 1.4. Finally, Section 1.5 concludes.

1.2 Birthplace diversity in the US states (1960-2010)

Following Ottaviano and Peri (2006), Ager and Brückner (2013), Suedekum et al. (2014) and Alesina et al. (2016), we consider that the cultural identity of individuals is mainly determined by their country of birth. The rationale is that the competitiveness of modern-day economies is closely linked to the average level of the human capital of workers and to the complementarity between their skills. On the one hand, workers originating from different countries were trained in different school systems and are more likely to bring complementary skills, cognitive abilities and productive traits. On the other hand, the diversity of the labor force can induce communication and cooperation costs. Whether or not the gains exceed the costs is a key empirical question.

In our empirical framework, our key explanatory variable is an index of birthplace diversity (or fractionalization), which can be computed for each US state and for the high-skilled and low-skilled populations separately. In subsection 1.2.1, we define various measures of birthplace diversity, establish links between them, and discuss their statistical correlation with the average immigration rate. In subsection 1.2.2, we then document the global US trends in cultural diversity observed since WW2.

[5]The GDP per capita of Hawaii (diversity index of 0.797) would be 11.66% larger if Hawaii had the same diversity index as the District of Columbia; the difference in high-skilled diversity explains about 4.7% of the total income gap between Hawaii and the District of Columbia in 2010.

1.2.1 The Birthplace Diversity Index

In line with existing studies, we first define a Herfindahl-Hirschmann index of birthplace diversity, a term which was first introduced by Alesina et al. (2016). This index is denoted by $TD^K_{s,t}$ and can be computed for the skill group $K = (L, H, A)$ (L for the low skilled, H for the high skilled, and A for both groups), for each state $s = (1, ..., S)$ and for each year $t = (1, ..., T)$. It measures the probability that two randomly-drawn individuals from the type-K population of a particular state originate from two different countries of birth. As shown by Alesina et al. (2016) in a cross-country setting, the birthplace diversity index is poorly correlated with genetic, linguistic or ethnolinguistic fractionalization indices. The index is defined as:

$$TD^K_{s,t} = \sum_{i=1}^{I} k^K_{i,s,t}(1 - k^K_{i,s,t}) = 1 - \sum_{i=1}^{I}(k^K_{i,s,t})^2, \tag{1.1}$$

where $k^K_{i,s,t}$ is the share of individuals of type K, born in country i, and living in state s, in the type-K resident population of the state at year t. Computing the birthplace diversity index requires collecting panel data on the structure of the population by region of destination, by country of origin, and by education level. Our sample includes all US states (as well as the District of Columbia) between 1960 and 2010 in ten-year intervals, i.e. $s = (1, ..., 51)$ and $t = (1960, ..., 2010)$. Our choice to conduct the analysis at the state level is guided by the availability of long-term data series on macroeconomic performance, and by the comparability with cross-country results. We identify a common set of 195 countries of origin, including the US as a whole.[6]

Building on Alesina et al. (2016), the additive decomposition of the diversity index allows to distinguish between the *Between* and the *Within* components of the diversity index, $TD^K_{s,t} = BD^K_{s,t} + WD^K_{s,t}$. On the one hand, the *Between* component $BD^K_{s,t}$ measures the probability that a randomly-drawn pair of type-K residents includes a native and an immigrant, irrespective of where the immigrant comes from:[7]

$$BD^K_{s,t} = 2k^K_{US,s,t}(1 - k^K_{US,s,t}).$$

On the other hand, the residual *Within* component $WD^K_{s,t}$ measures the probability that

[6] With the the exception of section 1.4.3, we disregard heterogeneity between US natives born in different states (e.g. a Texan native is considered identical to a Californian one). See subsection 1.2.2 for a detailed description of the data.

[7] In our specific case, $k^K_{US,s,t}$ represents the share of US natives in the type-K population of state s at time t.

a randomly-drawn pair of type-K residents includes two immigrants born in two different countries:

$$WD_{s,t}^{K} = \sum_{i \neq s}^{I} k_{i,s,t}^{K}(1 - k_{i,s,t}^{K} - k_{US,s,t}^{K}).$$

In the US context, the evolution of the birthplace diversity index among residents is almost totally driven by the change in the *Between* component of diversity, $BD_{s,t}^{K}$, which only depends on the proportion of immigrants. The median share of the *Between* component in total diversity, $BD_{s,t}^{A}/TD_{s,t}^{A}$, equals 98% and its quartiles are equal to 92% and 97%. Similar findings are found for the low-skilled and high-skilled populations. Consequently, birthplace diversity in group K is almost perfectly correlated with the state-wide proportion of immigrants.[8] On average, the Pearson correlation between $TD_{s,t}^{K}$ and the total share of immigrants in the population, $m_{s,t}^{K} = (1 - k_{US,s,t}^{K})$, equals 0.99 for all K. It is thus impossible to statistically disentangle the effects of diversity from those of the size of immigration. For this reason and in line with existing works, our empirical specification distinguishes between the size of immigration and the variety of immigrants.

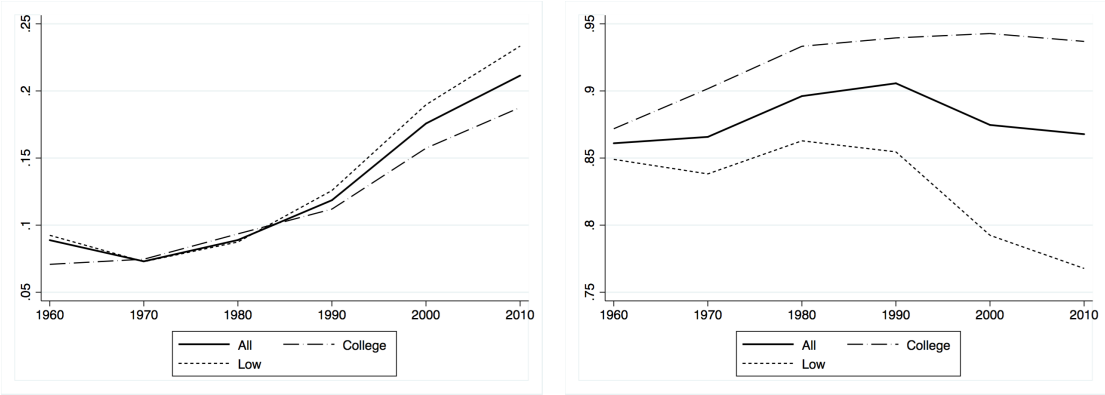
To capture the variety effect, we start from the *Within* component of the diversity index. The *Within* component can be expressed as the product of the square of the immigration rate (the probability that two randomly-drawn individuals are immigrants) by an index of diversity among immigrants, $MD_{s,t}^{K}$. The latter measures the probability that two randomly-drawn immigrants from state s originate from two different countries of birth. We have:

$$\begin{aligned} WD_{s,t}^{K} &= (1 - k_{US,s,t}^{K})^{2} MD_{s,t}^{K} \\ &= (1 - k_{US,s,t}^{K})^{2} \sum_{i \neq s} \widehat{k}_{i,s,t}^{K}(1 - \widehat{k}_{i,s,t}^{K}), \end{aligned} \tag{1.2}$$

where $\widehat{k}_{i,s,t}^{K} = k_{i,s,t}^{K}/(1 - k_{US,s,t}^{K})$ is the share of immigrants from origin country i in the total immigrant population of state s. Contrary to the total index of diversity and to its *Between* and *Within* components, the correlation between $MD_{s,t}^{K}$ and the total immigration rate, $m_{s,t}^{K}$, is small (on average, -0.19). This allows us to simultaneously include these two variables in the same regression without fearing collinearity problems.

[8] Table 1.A4 in the Appendix provides correlations between diversity indices, and between diversity and the immigration rate.

Figure 1.1: Trends in birthplace diversity in the US, 1960-2010



(a) Total diversity ($TD^K_{s,t}$)

(b) Diversity among immigrants ($MD^K_{s,t}$)

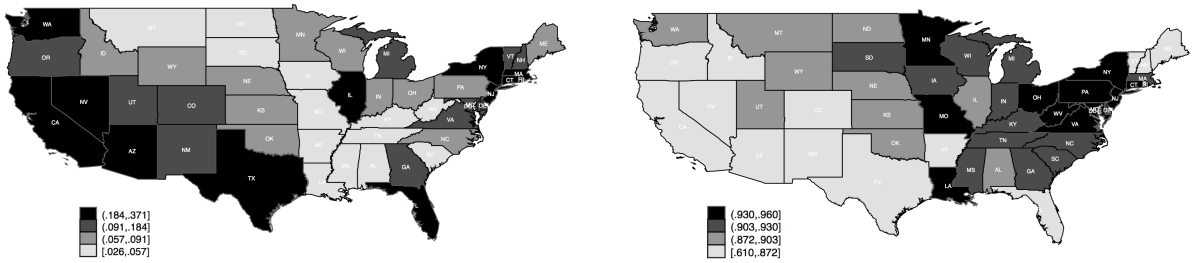
Notes: Total diversity (among residents) is defined as in Eq. (1.1), whereas diversity among immigrants is defined as in Eq. (1.2). Source: Authors' elaboration on IPUMS data.

1.2.2 Data and stylized facts

Population data at the state level for the US are available from the Integrated Public Use Microdata Series (IPUMS). IPUMS data are drawn from the federal census of the American Community Surveys. For each census year, they allow to characterize the evolution of the American population by country of birth, by age, by level of education, and by year of arrival in the US, among others. We extract the data from 1960 to 2010 in ten-year intervals, using the 1% census sample for the years 1960 and 1970, the 5% census sample for the years 1980, 1990 and 2000, and the American Community Survey (ACS-1%) sample for the year 2010. Regarding the origin countries of immigrants, we consider the full set of countries available in 2010, although some of them had no legal existence in the previous census years. Hence, for the years 1960 to 1990, data for the former USSR, former Yugoslavia and former Czechoslovakia are split using the country shares observed in the year 2000. In addition, we treat five pairs of countries as a single entity; this is the case of East and West Germany, Kosovo, Serbia and Montenegro, North and South Korea, North and South Yemen, and Sudan and South Sudan. Finally, we allocate individuals with a non-specified (or an imperfectly specified, respectively) country of birth proportionately to the country shares in the US population (or to the country shares in the US population originating from the reported region, respectively).

In our benchmark regressions, we restrict our micro sample to all individuals aged 16 to 64, who are likely to affect the macroeconomic performance of their state of residence. We distinguish between two skill groups. Individuals with at least one year of college are classified

Figure 1.2: Cross-state differences in birthplace diversity, 1960-2010 average index



(a) Diversity among residents ($TD^A_{s,t}$) (b) Diversity among immigrants ($MD^A_{s,t}$)

Notes: Diversity among residents is defined as in Eq. (1.1), whereas diversity among immigrants is defined as in Eq. (1.2). The two maps present the average birthplace diversity observed between 1960 and 2010. Alaska and Hawaii are not represented. Source: Authors' elaboration on IPUMS data.

as highly skilled, whereas the rest of the population is considered as low skilled. We define as US natives all individuals born in the US or in US-dependent territories such as American Samoa, Guam, Puerto Rico, the US Virgin Islands and other US possessions. Foreign-born individuals are referred to as immigrants. Observations are aggregated at state level using IPUMS sample weights.

In alternative regressions, we only consider immigrants who arrived in the US after a certain age, or immigrants who are likely to have a legal status. As for the age-of-entry correction, we sequentially eliminate immigrants who arrived before the age of 5, 6, ..., 25. In order to proxy the number of undocumented immigrants, we follow the "residual methodology" described in Borjas (2017), and use information on the respondents' characteristics (such as citizenship, working sector, occupation, whether they receive public assistance, etc.).

We use IPUMS data to identify the bilateral stocks and shares of international migrants, $k^K_{i,s,t}$, in the population of each state s, by country of origin i and by education level K in year t. We thus construct comprehensive matrices of "Origin $\times$ State $\times$ Skill" stocks and shares from 1960 to 2010 in ten-year intervals.[9] Missing observations are considered as zeroes, even if a positive number of immigrants is identified for an adjacent year.[10]

The evolution of the average index of cultural diversity is described in Figure 1.1, whereas Figure 1.2 represents differences in the average level of diversity across US states. Figure

[9]We distinguish between 195 countries of birth and 50 US states plus the District of Columbia. Countries and states are listed in Appendix 1.A1. Descriptive statistics by state are provided in Table 1.A3.

[10]We identify 33,145 origin-state dyads with zero migrants out of a sample of 59,670 observations (i.e. 55.5% of zeroes). The missing values are mostly concentrated in the years 1960 and 1970. Information on the country of origin is not reported for 0.3% of individuals at the micro level. We treat these observations as missing information when we aggregate our data by country of origin at state level.

1.1(a) describes the evolution of the birthplace diversity index computed for the resident population, $TD^K_{s,t}$ for $K = (L, H, A)$, between 1960 and 2010. Looking at the average of all US states, the birthplace diversity index among residents increased from about 0.09 in 1960 to 0.21 in 2010, reflecting the general rise in immigration to the US. A large portion of this change occurred after 1990. Nevertheless, this average trend conceals significant differences between US states and between skill groups. As far as cross-state differences are concerned, the number of immigrants drastically increased in states such as California (+195%) or New York (+91%); on the contrary, the number of foreign-born individuals remained small and stable in other states such as Montana or Maine. Regarding differences between skill groups, changes in immigration rates were larger for the low skilled than for college graduates, particularly after the year 1980. This is mainly due to the large inflows of low-skilled Mexicans observed over the last three decades, which drastically affected the level of diversity in states located on the West Coast and along the US-Mexican border (see Figure 1.2(a) below).

Second, Figure 1.1(b) describes the evolution of the diversity index computed for the immigrant population, $MD^K_{s,t}$ for $K = (L, H, A)$. It shows that the evolution of diversity in the immigrant population varies across skill groups. Diversity among college-educated immigrants has always been greater than diversity among the less educated. This might be due to the fact that college-educated migrants are less prone to concentrate in regions where large migration networks exist; they consider moving to more (geographically) diversified locations. Differences between skill groups drastically increased after 1960. On the one hand, diversity among high-skilled immigrants increased during the sixties and seventies, possibly due to the the Immigration and Nationality Act of 1965. Changes have been smaller since 1980 despite the Immigration Act of 1990, which allocated 50,000 additional visas (in the form of a lottery) to people from non-typical origin countries. On the other hand, diversity among low-skilled immigrants has fallen since 1980. Again, the latter decline is mainly explained by the large inflows of low-skilled Mexicans. Along the Mexican border and on the West Coast, the probability that two randomly-drawn immigrants were born in two different countries decreased as the share of Mexicans increased. This is also illustrated in Figure 1.2(b), which reveals significant cross-state differences in the average level of diversity among immigrants.

In sum, the evolution of diversity among immigrants varies across US states and over time. We show in the Appendix (see Figure 1.A3) that diversity among immigrants decreased in states located along the US-Mexican border and on the West Coast. A rise in diversity was observed in other states (such as Maine or Vermont). Our panel data analysis takes

advantage of these intra-state and inter-state variations to identify a causal effect of diversity on macroeconomic performance.

1.3 Empirical Strategy

Our goal is to assess the effect of birthplace diversity on the macroeconomic performance of US states.[11] The level of macroeconomic performance is measured by the log of the Gross Domestic Product (GDP) per capita. We present the benchmark specification in subsection 1.3.1. Subsection 1.3.2 explains how we deal with endogeneity issues, relying on placebo and IV regressions and combining two instrumentation strategies.

1.3.1 Benchmark Specification

Our benchmark empirical model features the log of GDP per capita as the dependent variable. In line with Ottaviano and Peri (2006), Ager and Brückner (2013), Suedekum et al. (2014) and Alesina et al. (2016), we use the following specification:

$$log(y_{s,t}) = \beta_1 MD^K_{s,t} + \beta_2 m^K_{s,t} + \boldsymbol{\lambda}' \boldsymbol{X}_{s,t} + \gamma_s + \gamma_t + \varepsilon_{s,t}, \tag{1.3}$$

where $log(y_{s,t})$ is the log of GDP per capita in state s at year t, $MD^K_{s,t}$ is the type-K birthplace diversity among immigrants (proxy for the variety of immigrants), and $m^K_{s,t}$ is the proportion of immigrants in the working-age population of type K. The latter is a key control variable that potentially captures the other channels through which the level of immigration affects macroeconomic performance (e.g. labor market, fiscal or market-size effects). We opt for a static specification and assume that changes in diversity fully materialize within 10 years. This spares us from dealing with the endogeneity of the lagged dependent, an important issue in dynamic models with a short-panel dimension (Nickel, 1981).[12] Coefficient β_1 is

[11] In the Appendix, a complementary analysis is conducted on the 34 OECD member states, using population data from Özden et al. (2011). The first drawback of the database is that it does not report the educational structure of migration stocks. To capture skill-specific effects, we combine it with the 1990-2000 estimates of the bilateral proportion of college graduates provided in Artuc et al. (2015). The second drawback is that it relies on imputation techniques to fill the missing bilateral cells. Despite the lower quality of the data, our fixed-effect analysis globally confirms the results obtained for US states.

[12] Nevertheless, Tables 1.A24 and 1.A25 in the Appendix provide the results of dynamic GMM regressions with internal or external instruments, and with different lag structures. In these regressions, the lagged dependent is insignificant or weakly significant, which reinforces the credibility of our static benchmark specification. In addition, the effects of diversity are qualitatively similar to those obtained in the static model.

our coefficient of interest. It captures the effect of birthplace diversity on macroeconomic performance.

Using skill-specific measures of cultural diversity and immigration, $K = (L, H, A)$, we can identify whether the level and significance of β_1 vary across skill groups. One might be concerned that using skill-specific regressions leads to an omitted variable bias. This can be the case if the origin mixes of high-skilled and low-skilled immigrants (and the skill-specific diversity indices) are correlated across states. However, the main reason for using skill-specific regressions is that the correlation between $m^H_{s,t}$ and $m^L_{s,t}$ exceeds 0.9. Hence, we estimate β_1 separately to avoid collinearity issues. Nonetheless, unreported regressions (available upon request) show that our results for diversity hold when including the two variables jointly.

We first estimate Eq. (1.3) using pooled OLS regressions, bearing in mind that such regressions raise a number of econometric issues that might generate inconsistent estimates. The key issue when using pooled OLS regressions is the endogeneity of the main variable of interest, the index of diversity. Endogeneity can be due to a number of reasons. These reasons include the existence of uncontrolled confounding variables causing both dependent and independent variables, the existence of a two-way causal relationship between these variables, or a measurement problem. To mitigate the possibility of an omitted variable bias, our specification includes a full set of state and year fixed effects, γ_s and γ_t, which allows us to partly account for unobserved heterogeneity (including initial conditions in 1960). In addition, our benchmark model includes a vector $\boldsymbol{X}_{s,t}$ of time-varying covariates through which immigration is likely to affect the level of macroeconomic performance. It includes the log of population, the log of the state-wide average educational attainment of the working-age population (as measured by the years of schooling or highest degree completed), and the log of the urbanization rate. As highlighted in Angrist and Pischke (2009), these control variables can be suspected of acting as "bad controls" if they are determined simultaneously with the level of diversity. For this reason, we provide supplementary OLS and FE regressions without control variables in section 1.4.1. They show that our results are highly robust to the inclusion or the exclusion of these controls.[13] Finally, to solve the reverse causation and measurement problems, we use placebo tests and two methods of instrumental variables described in subsection 1.3.2.

Finally, our benchmark specification in Eq. (1.3) assumes linear effects of the level of im-

[13]Table 1.A15 in the Appendix reports the results of IV regressions without controls. Remember that we also provide System-GMM regressions in Table 1.A24 and 1.A25 that correct for the endogeneity of each covariate using its own lags as instruments. These robustness checks confirm the results of our benchmark regressions.

migration and of the variety of immigrants on the log of GDP per capita. However, the literature on diversity suggests that the technology of transmission of cultural shocks can be different. In particular, another strand of the literature focuses on migration-induced transfers of norms, and tests for potential epidemiological or contamination effects. Transfers of norms from origin to destination countries have been examined by a limited set of studies.[14] Comparing the economic performance of US counties from 1850 to 2010, Fulford et al. (2017) show that the country-of-ancestry distribution of the population matters, and that the estimated effect of ancestry is governed by the sending country's level of economic development, as well as by measures of social capital at origin (such as trust and thrift). Putterman and Weil (2010) study the effect of ancestry in a cross-country setting, and find that the ancestry effect is governed by a measure of state centralization in 1500. More recently, debates about the societal implications of diversity have been revived in the migration literature. Collier (2013) and Borjas (2015) emphasize the social and cultural challenges that movements of people may induce.

The birthplace diversity index used in the benchmark specification neither accounts for characteristics of the origin countries not for the distance with the state of destination. Two strategies are used to account for origin-specific characteristics and potential epidemiological effects. The first consists in supplementing our specification with $MY^K_{s,t}$, the weighted average of the log of GDP per capita in the origin countries of type-K immigrants to state s (the weights are equal to the bilateral shares of immigrants). The epidemiological term is defined as:

$$MY^K_{s,t} = \sum_{i \neq s}^{I} \widehat{k}^K_{i,s,t} log(y_{i,t}). \tag{1.4}$$

On average, the correlation between this term and the diversity index is small (around -0.17 across US states), so that both variables can be tested jointly. Similarly, the correlation with the immigration rate is small as well (-0.26).[15]

[14] More studies focus on emigration-driven contagion effects, i.e. the effects of migrants' destination-country characteristics on outcomes at origin. The most popular study is that of Spilimbergo (2009), which investigates the effect of foreign education on democracy. Beine et al. (2013) and Bertoli and Marchetta (2015) use a similar specification to examine the effect of emigration on source-country fertility. Lodigiani and Salomone (2012) find that emigration to countries with greater female participation in parliament increases female participation in the origin country. More recently, Docquier et al. (2016) and Barsbai et al. (2017) find that emigration to democratic countries has a positive effect on the quality of institutions in developing countries.

[15] Alesina et al. (2016) control for such epidemiological terms and find insignificant effects. Compared to them, we consider several variants of Eq. (1.4) in the Appendix, and we also instrument epidemiological terms.

The second strategy consists in replacing $MD^K_{s,t}$ by an augmented Greenberg index (Greenberg, 1956) that accounts for the economic and cultural distances between the origin country and the US. By varying the parameters of the augmented index, we can uncover the conditions under which the beneficial or detrimental effects of diversity are maximized. In line with Alesina et al. (2016), our Greenberg diversity index is written:

$$MD^{K,Aug}_{s,t} = \sum_{i \neq s} \widehat{k}^K_{i,s,t}(1 - \widehat{k}^S_{i,s,t}) \times d_{US,i} \times e_{i,t} \tag{1.5}$$

The Greenberg index combines two measures of distance, $d_{US,i}$ and $e_{i,t}$ normalized between 0 and 1. For each distance we use a standard logistic function allowing us to weight these distances by factors θ_1 and θ_2 that vary between -10 and +10. The first distance ($d_{US,i}$) is the cultural distance between origin country i and the US:

$$d_{US,i} = 2 \Big/ \left(1 + e^{-(\theta_1 * D^d_{US,i})}\right) \tag{1.6}$$

For $D^d_{US,i}$, we alternatively use time-invariant measures of linguistic and genetic distance. Data on linguistic distance are taken from Head et al. (2010). $D^d_{US,i}$ is a dummy equal to one if a language is spoken by at least 9% of the population in both countries. Data on genetic distance are taken from Spolaore and Wacziarg (2009).[16]

The second distance ($e_{i,t}$) is a measure of economic distance between the origin country and the US:

$$e_{i,t} = 2 \Big/ \left(1 + e^{-(\theta_2 * D^e_{i,t})}\right) \tag{1.7}$$

Variable $D^e_{i,t}$ is proxied by the level of GDP per capita in the origin country only, to avoid using our dependent variable on the right-hand side of the regression equation. The use of panel data allows us to compute our Greenberg index with a time-varying measure of economic distance.

When both weights (θ_1 and θ_2) are equal to zero, our augmented index is equivalent to the standard Herfindahl index. However, when θ_1 increases (for a given θ_2), our Greenberg index over-weights immigrants from culturally distant countries. Similarly, when θ_2 increases (for

[16]Due to missing values, we eliminate 12 countries from our sample of 195 origins (Andorra, Bosnia, East Timor, Holy See, Liechtenstein, Marshall Islands, Micronesia, Monaco, Nauru, Occupied Palestinian Territories and Tuvalu). In 2010, these countries accounted for 0.3% of the high-skilled immigrant stock in the US.

a given θ_1), our Greenberg index over-weights immigrants from richer countries, i.e. from less economically distant countries.

1.3.2 Identification Strategy

Although our benchmark specification includes time-varying covariates and a full set of fixed effects, the association between birthplace diversity and macroeconomic performance can still be driven by reverse causality. As argued by Alesina and La Ferrara (2005), diversity is likely to respond to changes in the economic environment. The sign of the bias is ambiguous, especially when working at the state level. On the one hand, an upward bias can be observed if more rapidly growing states attract more immigrants from more diverse countries (e.g. due to higher income levels and business diversification). On the other hand, a downward bias can be observed if more rapidly growing states attract more immigrants from a limited set of origin countries (e.g. due to border and/or network effects). The example of states located on the Mexican border is highly relevant in this context (see Figures 1.2(a) and 1.2(b)).

Therefore, we use two strategies for dealing with reverse causation issues. As a placebo regression, we first augment our benchmark specification with natives' migration rates (denoted by $n^K_{s,t}$), and measures of diversity computed for the native population (denoted by $ND^K_{s,t}$). More precisely, we use the IPUMS data to identify the state of birth and the state of residence of each American citizen, and we compute internal migration rates and indices of diversity by state of birth for both skill groups. The latter index measures the probability that two randomly-drawn Americans from the type-K population of a particular state originate from two different states of birth. If diversity responds to economic prosperity, we expect a positive correlation between $ND^K_{s,t}$ and GDP per capita.

Secondly, we use two-stage least-square estimation methods. We consider two different sets of instruments that have been used in the existing literature, and show that our IV results are robust to the instrumentation strategy. Our first IV strategy is a shift-share strategy *a la* Ottaviano and Peri (2006) or Ager and Brückner (2013). We instrument diversity with predicted indices combining the changes in US immigration stocks by country of origin and the dyadic shares observed in 1960. More precisely, we predict the skill-specific bilateral migration stocks for each state using the residence shares of natives and immigrants observed in 1960. Then, we use these shares to allocate new immigrants by state of destination. The predicted stock of migrants at time t is:

$$\widehat{Stock}^K_{i,s,t} = Stock^K_{i,s,1960} + \phi^K_{i,s}(Stock^K_{i,t} - Stock^K_{i,1960}), \tag{1.8}$$

where $Stock^K_{i,s,t}$ is the type-K stock of immigrants from country i residing in state s at year t. Term $\phi^K_{i,s}$ is the time-invariant share that we use to allocate the variation in the bilateral migration stocks observed between the years 1960 and t. We thus allocate changes in bilateral migration stocks using the 1960 skill-specific shares of US natives and immigrants from the same origin country. These shares capture both origin- and skill-specific network effects, and the concentration of type-K workers in 1960. We have:

$$\phi^K_{i,s} = \Big(Nat^K_{s,1960} + Stock^K_{i,s,1960}\Big) \Big/ \Big(\sum_s (Nat^K_{s,1960} + Stock^K_{i,s,1960})\Big), \tag{1.9}$$

where $Nat^K_{s,1960}$ is the number of US natives residing in state s at year 1960. Using the predicted stock of migrants (who are less likely to be affected by the economic performance of each state), we compute the predicted diversity indices.

In line with Feyrer (2019) or Alesina et al. (2016), our second IV strategy consists in instrumenting diversity indices using the predicted migration stocks obtained from a "zero-stage," pseudo-gravity regression. The latter regression includes interactions between year dummies and the geographic distance between each country of origin and each US state. In line with the shift-share strategy, the identification thus comes from the time-varying effect of geographic distance on migration, reflecting gradual changes in transportation and communication costs. The pseudo-gravity model is written:

$$\log(Stock^K_{i,s,t}) = \beta_t \log(Dist_{i,s}) + Bord_{i,s} + Lang_{i,s} + \gamma_s + \gamma_i + \gamma_t + \varepsilon_{i,s,t}, \tag{1.10}$$

where $Bord_{i,s}$ is a dummy equal to one if country i and state s share a common border, $Lang_{i,s}$ is a dummy equal to one if at least 9% of the populations of i and s speak a common language, γ_s, γ_i, and γ_t are the destination, origin and year fixed effects.[17] In the pseudo-gravity stage, the high prevalence of zero values in bilateral migration stocks gives rise to econometric concerns about possible inconsistent OLS estimates. To address this problem, we use the Poisson regression by pseudo-maximum likelihood (see Santos Silva and Tenreyro, 2006). Standard errors are robust and clustered by country-state pairs.

Although commonly used in the literature, each of these IV strategies has some drawbacks. We are aware that the shift-share method is imperfect if potential regressors exhibit strong persistence, and that the relative geography variables used in the strategy *a la* Feyrer (2019) can affect macroeconomic performance through other channels such as trade, foreign direct

[17] One could be concerned that the border dummies in the pseudo-gravity equation impose a very strong functional form input on the instrumental variables estimates. However, it is worth noting that removing these two variables from our equation does not change significantly our main coefficients of interests.

investments or technology diffusion (not measurable at the state level for the 1960-2010 period). Nevertheless, we can reasonably support a careful causal interpretation of our results if these strategies and the placebo regressions yield consistent and converging results.

1.4 Results

Our empirical analysis follows the structure described in section 1.3. In subsection 1.4.1, we investigate the effect of birthplace diversity among immigrants using pooled OLS regressions; we produce separate results for the total immigrant population and for the two skill groups. Then, we test for the existence of epidemiological effects, and we control for unobserved heterogeneity, including a full set of state and year fixed effects (FE). In subsection 1.4.2, we show that the FE estimates are highly robust to different robustness checks and subsamples. In subsection 1.4.3, we deal with endogeneity issues, discussing the results of placebo and IV regressions. The latter rely on two instrumentation strategies frequently used in the existing literature (i.e a shift-share strategy *a la* Ottaviano and Peri (2006) and a gravity-like strategy *a la* Feyrer (2019)). Finally, in subsection 1.4.4, we extend the IV setting and conduct an exploratory analysis of the channels of transmission through which birthplace diversity affects economic growth.

1.4.1 Pooled OLS and FE Regressions

Table 1.1 describes the pooled OLS and FE estimates without and with covariates in Tables 1.1(a) and 1.1(b), respectively. We produce separate results for the three skill groups, $K = (A, L, H)$, under the same set of potential confounding factors. These include the skill-specific immigration rate, $m^K_{s,t}$, the log of population, $log(Pop_{s,t})$, the log of urbanization, $log(Urb_{s,t})$, and the log of the average educational attainment of the working-age population, $log(Hum_{s,t})$. In all cases, our standard errors are clustered at the state level in order to correct for heteroskedasticity and serial correlation.

The pooled OLS estimates are reported in col. 1, 3 and 6. The effect of birthplace diversity on GDP per capita is skill-specific. Insignificant effects are obtained when diversity is computed using the low-skilled immigrant population. On the contrary, the association between GDP per capita and high-skilled diversity is positive and significant at the 1% level. The coefficient is large: a 10% increase in high-skilled diversity is associated with a 27.2% increase in GDP per capita when potential confounding factors are accounted for, and even

Table 1.1: Pooled OLS and FE regressions
Results by skill group (Dep= $log(y_{s,t})$)

	(1) OLS	(2) FE	(3) OLS	(4) FE	(5) FE	(6) OLS	(7) FE	(8) FE
Skill level	All	All	Low skilled	Low skilled	Low skilled	High skilled	High skilled	High skilled
Table 1.1(a): Without control variables								
$Mig\ Div_{s,t}$	2.084***	0.275***	0.343	0.147**	0.096	7.595***	0.664***	0.556***
	(0.581)	(0.097)	(0.498)	(0.063)	(0.060)	(0.909)	(0.209)	(0.182)
$Mig\ Share_{s,t}$	8.905***	-0.115	6.963***	-0.123	-0.152	9.912***	-0.019	-0.265
	(1.267)	(0.402)	(1.053)	(0.314)	(0.315)	(1.421)	(0.452)	(0.452)
$MY_{s,t}$					-0.150***			-0.174**
					(0.048)			(0.068)
Constant	7.099***	7.847***	8.738***	7.960***	9.285***	1.931**	7.501***	9.095***
	(0.523)	(0.087)	(0.436)	(0.059)	(0.419)	(0.861)	(0.187)	(0.614)
R-squared	0.283	0.991	0.236	0.991	0.992	0.452	0.992	0.992
Table 1.1(b): With control variables								
$Mig\ Div_{s,t}$	0.416	0.318***	0.019	0.141	0.104	2.719***	0.616***	0.531***
	(0.329)	(0.114)	(0.184)	(0.086)	(0.085)	(0.719)	(0.160)	(0.159)
$Mig\ Share_{s,t}$	2.632***	0.582*	1.901***	0.481*	0.412	4.383***	0.614*	0.388
	(0.615)	(0.341)	(0.485)	(0.282)	(0.283)	(1.018)	(0.315)	(0.366)
$MY_{s,t}$					-0.104**			-0.133*
					(0.042)			(0.069)
$log(Pop_{s,t})$	0.070	-0.172**	0.079*	-0.166**	-0.146*	0.011	-0.155**	-0.080
	(0.047)	(0.079)	(0.047)	(0.081)	(0.082)	(0.044)	(0.075)	(0.065)
$log(Urb_{s,t})$	-0.407*	0.385**	-0.367	0.329**	0.312*	-0.563**	0.285**	0.156
	(0.238)	(0.156)	(0.254)	(0.163)	(0.173)	(0.229)	(0.135)	(0.138)
$log(Hum_{s,t})$	5.752***	0.695***	5.817***	0.807***	0.802***	5.288***	0.759***	1.007***
	(0.157)	(0.197)	(0.147)	(0.205)	(0.196)	(0.182)	(0.197)	(0.299)
Constant	-0.697	7.529***	-0.728	7.662***	8.379***	-0.584	7.348***	7.492***
	(0.890)	(1.254)	(0.914)	(1.263)	(1.317)	(0.890)	(1.262)	(1.273)
R-squared	0.879	0.993	0.878	0.993	0.993	0.889	0.993	0.993
Observations	306	306	306	306	306	306	306	306
Nb. states	51	51	51	51	51	51	51	51
Time fixed effects	No	Yes	No	Yes	Yes	No	Yes	Yes
States fixed effects	No	Yes	No	Yes	Yes	No	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. The specification is described in Eq. (1.3). Pooled OLS results are provided in col. 1, 3 and 6; FE results are provided in col. 2, 4, 5, 7 and 8. Results for all immigrants are provided in col. 1 and 2; results for low-skilled immigrants are provided in col. 3, 4 and 5; results for college-educated immigrants are provided in col. 6, 7 and 8. The sample includes the 50 US states and the District of Columbia from 1960 to 2010. The set of control variables includes the immigration rate ($Mig\ Share_{s,t}$), the log of population ($log(Pop_{s,t})$), the log of urbanization ($log(Urb_{s,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{s,t})$). We supplement our benchmark specification in col. 5 and 8 with the epidemiological effect ($MY_{s,t}$).

more without controls. When computing the diversity index on the total immigrant population, the effect remains significant at the 1% level in the regression without controls, but it becomes insignificant when the covariates are included.

Col. 2, 4 and 7 describe the (benchmark) results obtained when we add state and year fixed effects. The state fixed effects account for all time-invariant state characteristics that could jointly affect productivity and diversity; the year fixed effects account for time-varying sources of change in GDP per capita that are common to all US states.[18] In the FE regressions, the R-squared is above 0.99. The effect of high-skilled diversity remains highly significant. Interestingly, the inclusion of fixed effects leads to a drop in the coefficient of interest. It is divided by four compared to the pooled OLS regression. This demonstrates that accounting for unobserved heterogeneity is crucial in this literature. As far as low-skilled diversity is concerned, the effect becomes significant in the regression without controls, but loses significance once covariates are factored in. Turning our attention to the potential confounding factors, human capital and urbanization rates are significantly and positively associated with GDP per capita. On the contrary, the correlation between GDP and population size is negative. More interestingly, immigration rates are always positively associated with GDP per capita, and the correlation is always greater for college graduates.

In col. 5 and 8, we supplement the benchmark model with epidemiological effects *a la* Collier (2013) and Borjas (2015). Interpreting the coefficient of the epidemiological term is not straightforward. On the one hand, if immigrants originating from poor countries contaminate the total factor productivity or the quality of institutions at destination, we should find a positive and significant relationship between our epidemiological term ($MY^K_{s,t}$) and macroeconomic performance. On the other hand, if attracting immigrants from economically or culturally distant countries generates more complementarities in skills and ideas than immigrants from richer countries, we should find a negative and significant relationship. Overall, controlling for epidemiological terms does not affect our findings. The effect of low-skilled diversity remains insignificant in col. 5, and the effect of high-skilled diversity remains highly significant in col. 8. These results are robust to the specification of the epidemiological terms.[19] In addition, it is worth noticing that we find no evidence of contamination effects after controlling for potential confounding factors. In col. 8, the coefficient of the epidemi-

[18] In order to better deal with unobserved heterogeneity, additional regressions with region-by-year fixed effects are provided in the Appendix. As shown in Table 1.A17, our findings are not altered.

[19] We consider alternative specifications for the epidemiological term in Appendix Table 1.A19. We first compute $MY^K_{s,t}$ by keeping the immigration shares ($\widehat{k}^K_{i,s,t}$) constant, at their 1960-2010 average levels. Then, we keep the levels of GDP per capita at origin ($log(y_{i,t})$) constant, at their 1960-2010 average level. Finally, we combine annual data on GDP per capita at origin with individual data on the year of arrival in the US; each immigration share is multiplied by the average level of GDP per capita prevailing in the year of immigration to the US. The latter specification allows us to capture the norms and values that immigrants bring with them when they migrate. Due to data limitations, this variable cannot be computed for the year 1960. Finally, we obtain the same conclusion when the epidemiological term in Eq. (1.4) is based on democracy levels at origin, instead of GDP per capita. We use the Polity2 index of democracy. The results for democracy are available upon request.

ological terms is weakly significant and negative. In col. 5, the coefficient is negative and significant at the 5% level. Such a negative effect suggests that low-skilled immigrants from richer countries may generate fewer complementarities with the native workforce compared to immigrants from poorer countries. However, reverse causality is a source of concern here since economic growth is likely to affect the attractiveness of a state and the variety of its immigrant population. Our database reveals that richer states attract more people, including immigrants from poorer countries. This selection issue pushes the correlation between GDP per capita and the epidemiological term downwards.[20] Such reverse causation issues are addressed in subsection 1.4.3.

In sum, we find that diversity is positively associated with the level of GDP per capita, but only when diversity is computed on workers performing complex or skill-intensive tasks. On the contrary, diversity among less educated immigrants does not have a significant effect on macroeconomic performance. According to our fixed-effect estimates, a 10% increase in high-skilled diversity (i.e. in the probability that two randomly-drawn, college-educated immigrants originate from two different countries of birth) is now associated with a 6.2% increase in GDP per capita. Expressed differently, a one-standard-deviation increase in high-skilled diversity is associated with a 3.2% increase in GDP per capita. This implies that, if all US states had the same level of diversity as the most diverse state in 2010, i.e. the District of Columbia (0.976), the average GDP per capita of the US would be 2.33% larger, the coefficient of variation across states would be 2.37% smaller, and the Theil index would decrease by 3.45%. By comparison, if all US states had the same average level of human capital as the District of Columbia, the average GDP per capita of the US would be 8.32% larger, the coefficient of variation across states would be 9.77% smaller, the Theil index would decrease by 16.06% and the GDP per capita of Hawaii, the least diverse state in 2010 (0.797), would be 11.7% larger. In addition, the US-state average level of diversity among college-educated migrants increased by 7 percentage points between 1960 and 2010; this explains a 3.5% increase in macroeconomic performance (i.e. only one fiftieth of the total change in the US level of GDP per capita).

1.4.2 Robustness checks

This subsection investigates the robustness of the previous results. Table 1.2 summarizes the results for high-skilled diversity. With the exception of the analysis by Commuting Zone

[20]Figure 1.A4 in the Appendix confirms this presumption. When we keep the levels of GDP per capita constant for all origin countries (at their 1960-2010 average), we observe that the US state level of GDP per capita is negatively correlated with the epidemiological term.

Table 1.2: Robustness of FE regressions for high-skilled diversity
(Dep= $log(y_{s,t})$)

	Sub-sample (1.A7)						CZ (1.A9)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Full-Sample	1970-2000	No Top5	No Bot5	No Mex	No Q1	*CZ*
$Mig\ Div_{s,t}$	0.616***	0.870***	0.725***	0.672***	0.630***	0.596**	0.319**
	(0.160)	(0.321)	(0.174)	(0.170)	(0.170)	(0.288)	(0.150)
$Mig\ Share_{s,t}$	0.614*	1.140**	1.317**	0.613*	0.541	0.765**	1.907***
	(0.315)	(0.459)	(0.529)	(0.323)	(0.397)	(0.365)	(0.567)
Observations	306	204	276	276	282	228	3,688
Nb. states/CZs	51	51	46	46	47	38	741
R-squared	0.993	0.990	0.993	0.993	0.993	0.995	0.895

	10 Largest (1.A10)	Quadratic (1.A12)	Educ. level (1.A13)		Legal status (1.A14)	
	(8)	(9)	(10)	(11)	(12)	(13)
			Ph.D	Tertiary	Docum.	Undoc.
$Mig\ Div_{s,t}$	0.617***	-0.131	0.262**	0.369***	1.009**	-0.153
	(0.169)	(1.954)	(0.103)	(0.136)	(0.473)	(0.127)
$Mig\ Share_{s,t}$	0.726*	0.622*	0.256	0.372	0.959*	4.426**
	(0.365)	(0.314)	(0.266)	(0.298)	(0.535)	(2.140)
$(Mig\ Div_{s,t})^2$		0.453				
		(1.202)				
Observations	306	306	306	306	204	204
Nb. states	51	51	51	51	51	51
R-squared	0.993	0.993	0.993	0.993	0.979	0.979

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state (or CZ) level. A-indexed numbers in parentheses refer to full Tables provided in the Appendix. All models (except col. 7) include the full vector of controls (not shown) with the log of population ($log(Pop_{s,t})$), the log of urbanization ($log(Urb_{s,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{s,t})$) as well as time and state fixed effects. ⋆ The sample for col. 7 includes the geographical area of 50 US states and the District of Columbia from 1970 to 2010. The dependent variable is the logarithm of the average wage of white US natives between 40-50.

(CZ), all regressions include the full set of time and state fixed effects and the set of covariates used in Table 1.1(b). The estimates for the covariates are not reported here; complete tables are provided in the Appendix. In addition, robustness checks for low-skilled diversity are reported in Table 1.A6 in the Appendix. Irrespective of the specification we always find insignificant effects of low-skilled diversity on GDP per capita.

Robustness by subsample. – In Table 1.2 , the benchmark results of Table 1.1(b) are reported in col. 1. In col. 2, we limit our sample to the 1970-2000 period, eliminating possible sources of variation prior to the 1965 amendments to the Immigration and Nationality Act, as well as variations driven by the recent evolution of diversity. Remember that Figure 1.1(b) shows that the average high-skilled diversity index slightly decreased between 2000 and 2010.

Then, in col. 3 and 4, we examine whether the impact of diversity is driven by the size of the immigrant population: we drop the five US states with the greatest or the smallest immigration rates in 2010, respectively.[21] In col. 5, we investigate whether our results are driven by the Mexican diaspora, which represented 30% of the whole immigrant population of the US in 2010. We drop the states located on the US-Mexican border, which host 62% of all Mexican immigrants to the US.[22] Remember that these states have experienced a drastic decrease in their diversity index (-40% in low-skilled diversity between 1960 and 2010), which is entirely due to the rising inflows from Mexico.[23] Finally, in col. 6, we exclude the states in the first quartile (i.e. below Q1) of the 2010 distribution by immigrant population size.

Overall, we show that our FE results are robust to sample selection. In Table 1.2, the coefficient of high-skilled diversity is always positive, significant, and of the same order of magnitude as the benchmark estimates in col. 1. The positive impact becomes even larger when reducing the time span (0.87) or after excluding the states with the highest immigration rates (0.73). This suggests that high-skilled diversity could generate non-linear effects on macroeconomic performance; we will explore this hypothesis in col. 9 as well as in subsequent sections.

Robustness to the geographical unit of analysis. – In col. 7, we check whether our results hold when measuring diversity and immigration at the Commuting Zone (CZ) level, which better captures local labor market conditions. We use the CZ classification described in Dorn (2009). Since GDP per capita is not measurable at the CZ level, we follow Ottaviano and Peri (2006) and compute the logarithm of the average weekly earnings of white US natives aged 40 to 50 as a dependent variable. Unfortunately, using weekly earning forces us to restrict our sample to the 1970-2000 period due to data constraints on the number of weeks worked in the 1960 census. Remember the decades 1960s and 1970s include the years with the greatest sources of variation in birthplace diversity. Our results at the CZ level are qualitatively similar to those obtained at the state level. We find a positive and significant effect of high-skilled diversity on the weekly wages of natives although it becomes smaller and less significant.

[21] The states with the greatest immigration rates are California, New York, Hawaii, New Jersey, and Florida. The states with the smallest rates are West Virginia, Mississippi, Kentucky, South Dakota, and Alabama.

[22] They include California, Texas, New Mexico, and Arizona.

[23] As an additional robustness check, we also drop the Mexican diaspora from the computation of the birthplace diversity index and the immigration rate, or we add the share of Mexican immigrants in the state as a separate variable. This increases the coefficients of high-skilled and low-skilled diversity, the latter becoming significant at the 5% level. Expanding the group of Mexican immigrants to all Hispanic immigrants gives the same results. See Table 1.A12 in the Appendix.

Controlling for large groups. – We also investigate whether the effect of birthplace diversity is not driven by the presence of large diasporas characterized by specific productivity levels (this generalizes what we did when excluding states located on the US-Mexican border). In col. 8, we control for the state-specific shares of the ten largest origin countries in the US immigrant population. Controlling for the size of the largest immigrant groups affects neither the significance nor the magnitude of our coefficient of interest.

Quadratic specification. – When estimating the effect of genetic diversity on economic development, Ashraf and Galor (2013) and Ashraf et al. (2015) consider a quadratic specification, which allows them to identify an optimal level of genetic diversity. In our context, diversity may also induce costs and benefits, implying that its effect on macroeconomic performance could be better captured by an inverted-U shaped relationship. In col. 9, we thus supplement our benchmark specification by adding the square of the diversity index. If an optimal level of diversity among immigrants exists, we should find a positive coefficient for the linear term, and a negative coefficient for the squared term. We find no evidence of a quadratic effect of birthplace diversity. The coefficient for the squared index of diversity is insignificant in Table 1.2. Hence, this regression rejects the existence of an optimal level of diversity among college-educated immigrants.

Robustness by skill group. – One might be concerned that the positive effect of high-skilled diversity is driven by the presence of immigrants at the very top of the skill distribution. We investigate this issue in col. 10 and 11 of Table 1.2. We find insignificant differences when computing diversity on PhD graduates, or on other college-educated immigrants.

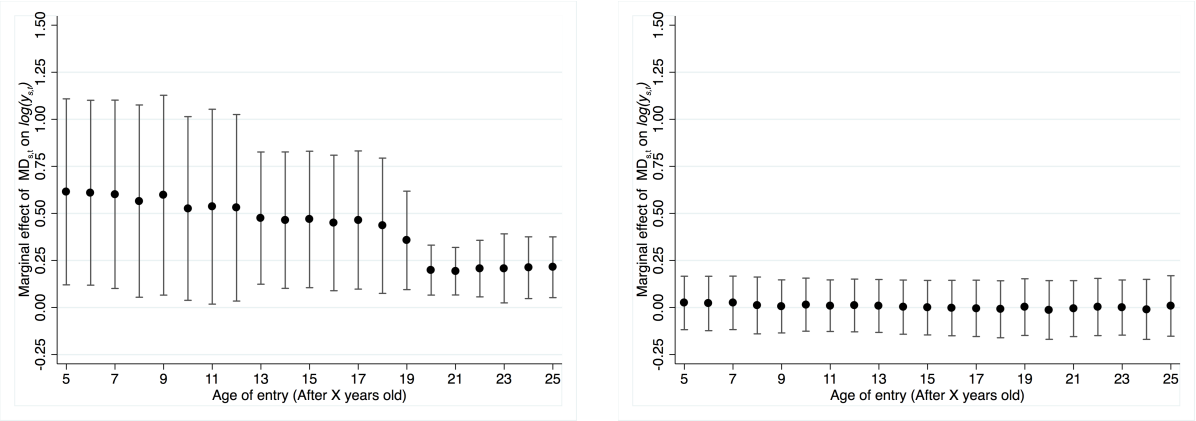
Robustness by legal status. – We also investigate the role of undocumented migration in governing the skill-specific effects of diversity. The US census counts every individual regardless of immigration status. Hence, undocumented immigrants influence our diversity index. This can be a source of concern as they are likely to be less educated than legal immigrants and to contribute differently to GDP, either because their productive activities are not recorded in the official GDP or because they are employed in jobs/sectors where skill complementarities are smaller. This could explain why the effect of low-skilled diversity is insignificant in most of our regressions. To explore this hypothesis, we use the "residual methodology" proposed by Borjas (2017) to identify the number of legal and undocumented immigrants by skill group. It consists in using individual characteristics to proxy the legal status of US immigrants. In this work, we use five characteristics (citizenship, employment industry, occupation, whether the individual receives any assistance, and the spouse's legal status) and, due to data availability, we apply the residual methodology to the census years 1980

to 2010. We thus compute the diversity indices on the legal and undocumented immigrant populations, and include them separately in our FE regressions in Col. 12 and 13 in Table 1.2. Distinguishing between legal and undocumented immigrants yields different effects. Diversity among undocumented immigrants has no significant effect, while diversity among legal immigrants has a positive and significant effect at the five% level. On the contrary, controlling for the legal status of low-skilled immigrants does not modify our conclusions (see Table 1.A6 in the Appendix). This confirms that the insignificant effect of low-skilled diversity cannot be attributed to the greater proportion of undocumented migrants in this group (on average, 17% for the US in 2010).

Robustness by age of entry. – The diversity indices used in our benchmark regressions are computed for the total population of working-age immigrants, whatever their age of entry in the US. As birthplace diversity conceivably reflects complementarities between individuals trained in different countries, it can be argued that immigrants who arrived in the US at different ages generate different levels of complementarity in skills and ideas with the native workforce. However, the role of the age of entry is unclear. On the one hand, immigrants with a longer foreign education are likely to bring more complementarities. On the other hand, immigrants who were partly educated in the US may have more transferable skills and a greater potential to interact with natives. To investigate this issue, we compute the diversity index using various samples of immigrants, and we include these alternative indices in Eq. (1.3). More precisely, we exclude from the immigrant population the individuals who arrived in the US before a given age threshold, which ranges from 5 to 25 in one-year intervals. For each skill group, Figure 1.3 reports the marginal effect of diversity and its confidence interval as a function of the age-of-entry threshold.[24] As information on age of entry is not available in the 1960 census, our sample covers the 1970-2010 period. For this time span, the coefficients of the benchmark FE regressions (without controlling for age of entry) are equal to 0.835 for high-skilled diversity (significant at the 1% level), and to 0.088 for low-skilled diversity (insignificant). Whatever the age-of-entry threshold, the effect of low-skilled diversity is insignificant. Nevertheless, the age of entry matters for college graduates. Although the coefficient of high-skilled diversity is always positive and significant, the largest effects are obtained when the immigrant population includes individuals who arrived before age 20. Considering three age thresholds (12, 18, and 22), Alesina et al. (2016) show that the positive effect of birthplace diversity slightly decreases when eliminating immigrant children, but always remains large and significant. Conversely, our results suggest that the greatest levels of complementarity are obtained when immigrants acquired primary and part of their

[24]Comprehensive regression results are provided in Table 1.A16 in the Appendix.

Figure 1.3: Marginal effect of $Mig\ Div_{s,t}$ on $log(y_{s,t})$
Results for different age-of-entry thresholds (1970-2010)



(a) High-skilled (b) Low-skilled

Notes: The two graphs report the marginal effect of $Mig\ Div_{s,t}$ on $log(y_{s,t})$ when the immigrant population is restricted to individuals who arrived in the US after age X. Marginal effects are obtained using our main specification Eq. (1.3) which includes state and year fixed effects, as well as the immigration rate ($Mig\ Share_{s,t}$), the log of population ($log(Pop_{s,t})$), the log of urbanization ($log(Urb_{s,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{s,t})$). Source: Authors' elaboration on IPUMS data.

secondary education abroad, and their college education in the US.

Additional robustness checks. – Many other robustness checks were carried out to validate our results and reinforce their interpretation. Some results can be found in the Appendix; others are available upon request. Firstly, one might be concerned that our results are driven by other dimensions of diversity that could be correlated with the variety of immigrants. In unreported regressions, we controlled for the racial or linguistic diversity within the native population. The results are unaltered.

Secondly, we also computed the diversity index after aggregating immigrants by continent or by broad region of origin. The results are available in Table 1.A18 in the Appendix. Birthplace diversity then becomes insignificant. This reinforces our presumption that the effect of diversity captures skill complementarities that are driven by the heterogeneity in origin country characteristics. These complementarities are poorly captured when aggregating countries within a broad region.

Thirdly, we checked whether our results hold when using alternative measures of diversity. Table 1.A12 in the Appendix shows that they hold when replacing our Herfindahl index by a Theil index of diversity. The advantage of the Theil index is that it can be decom-

posed into two additive components: a within-origin component that captures changes in the concentration of immigrants at the intensive margin, and a between-origin component that captures changes in the concentration of immigrants at the extensive margin (see (Cadot *et al.*, 2011)). Using this decomposition, we show that the positive effect of high-skilled diversity is driven by the intensive margin (i.e. by a more balanced mix of origin countries with positive stocks rather than by the entry of immigrants from new origin countries). In the same vein, our results hold when using a measure of cultural polarization. In Ager and Brückner (2013), Montalvo and Reynal-Querol (2003) and Montalvo and Reynal-Querol (2005), the index of polarization captures how far the distribution of a population is from the bimodal distribution. It is defined as:

$$TP_{s,t}^{K} = 1 - \sum_{i=1}^{I} \left((0.5 - k_{i,s,t}^{K})/0.5\right)^{2} k_{i,s,t}^{K} \tag{1.11}$$

Applied to the immigrant population (i.e. using $\widehat{k}_{i,s,t}^{K}$ instead of $k_{i,s,t}^{K}$ in the previous equation), the index $MP_{s,t}^{K}$ is maximized when there are two groups of immigrants which are of equal size (i.e. 50%). For US states, the polarization index exhibits a correlation of -0.89 with the diversity index (this is much greater than the correlation reported in Ager and Brückner (2013)). Hence, including these two variables in the same regression is risky. As shown in Table 1.A12 in the Appendix, replacing $MD_{s,t}^{K}$ by $MP_{s,t}^{K}$ gives the same results as in our benchmark regression but with the opposite sign.

Fourthly, we investigated whether the same effects of diversity are obtained when considering alternative proxies for macroeconomic performance. In particular, population and employment levels may also react to changes in total factor productivity. In unreported regressions, we did not find any significant effect of diversity on the log of employment or population as a dependent variable. This might be due to the fact that, in a context with mobility costs and with two types of workers, the relationship between productivity growth and aggregate population/employment growth is less mechanical than in a context with perfect labor mobility and one class of workers. Our results do not support an influence of diversity beyond the income effect.

Finally, we tested whether the size of skill complementarities between diverse immigrants can be attenuated by the cultural distance between them. Our birthplace diversity index $MD_{s,t}^{K}$ does not account for the cultural distance between origin and destination countries. It assumes that all groups are culturally equidistant from each other. We considered an adjusted index that multiplies the probability that two randomly-drawn immigrants were born in two different countries by a measure of cultural distance between these two countries. For the

latter, we used the database on genetic distance of Spolaore and Wacziarg (2009). Genetic distance is based on blood samples and proxies the time since two populations had common ancestors. Again, the results reported in 1.6.6 are robust to the use of an adjusted diversity index.

1.4.3 Dealing with endogeneity

In this section, we investigate the likelihood that our results are driven by reverse causality. We use placebo regressions in sub-section 1.4.3 and IV regressions in sub-section 1.4.3, treating birthplace diversity, immigration rates and epidemiological terms as potentially endogenous variables.[25]

Placebo regressions

We now investigate whether our results can be driven by reverse causality (i.e. growing states attracting more diverse immigrants). If diversity increases with economic prosperity, we can expect a positive correlation between birthplace diversity among American workers and GDP per capita, as explained in Section 1.3. In our placebo regressions, we augment the benchmark model with two additional control variables, namely the natives' migration rates ($n^K_{s,t}$) and the measures of diversity computed for the native population ($ND^K_{s,t}$). Since the range of variation of the diversity indices for immigrants and natives are different, we standardize $ND^K_{s,t}$ in such a way that it exhibits the same mean and standard deviation as $MD^K_{s,t}$.

Table 1.3 reports the results of our placebo regressions: internal immigration rates are positively correlated with GDP per capita. However, the native diversity index is insignificant (or weakly significant in col. 3). Although these placebo tests do not necessarily imply that diversity among foreign immigrants is not affected by macroeconomic performance, they mitigate the risk of a strong reverse causation relationship.

[25] Another issue relates to self-selection if immigrants from different source countries have unobserved characteristics and traits that are conducive to economic growth and that are captured by the diversity index. We deal with this omitted variable issue in Table 1.A23 in the Appendix. First, we follow Alesina et al. (2016) and assume that self-selection on unobserved productive traits can be proxied by self-selection on observable characteristics. We construct an index of selection on educational attainment in that we include it in the benchmark specification. Still, it could be possible that self-selection on unobserved traits is poorly captured by self-selection on educational attainment. Thus, we extend the analysis above and follow the methodology proposed by Oster (2019) on the likelihood of an omitted variable problem due to self-selection on unobservables. In both cases we find that self-selection is unlikely to drive our results.

Table 1.3: $Mig\ Div_{s,t}$ v.s diversity among "native immigrants" $Nat\ Div_{s,t}$
Results by skill group (Dep= $log(y_{s,t})$)

	(1)	(2)	(3)	(4)	(5)	(6)
Skill level	High skilled	High skilled	High skilled	Low skilled	Low skilled	Low skilled
$Mig\ Div_{s,t}$	0.616***		0.432**	0.141		0.156*
	(0.160)		(0.168)	(0.086)		(0.082)
$Mig\ Share_{s,t}$	0.614*		1.008***	0.481*		0.642**
	(0.315)		(0.347)	(0.282)		(0.278)
$Nat\ Div_{s,t}$		0.719	0.879*		0.139	0.157
		(0.486)	(0.492)		(0.174)	(0.172)
$Nat\ Share_{s,t}$		0.376**	0.428**		0.059	0.218
		(0.167)	(0.203)		(0.242)	(0.244)
$log(Pop_{s,t})$	-0.155**	-0.135*	-0.176**	-0.166**	-0.120*	-0.169**
	(0.075)	(0.068)	(0.073)	(0.081)	(0.067)	(0.073)
$log(Urb_{s,t})$	0.285**	0.294*	0.316**	0.329**	0.266	0.304*
	(0.135)	(0.159)	(0.149)	(0.163)	(0.164)	(0.172)
$log(Hum_{s,t})$	0.759***	0.557**	0.477**	0.807***	0.677**	0.505*
	(0.197)	(0.213)	(0.217)	(0.205)	(0.271)	(0.269)
Constant	7.348***	7.058***	7.091***	7.662***	7.493***	8.049***
	(1.262)	(1.377)	(1.389)	(1.263)	(0.988)	(1.027)
Observations	306	306	306	306	306	306
Nb. states	51	51	51	51	51	51
R-squared	0.993	0.993	0.993	0.993	0.993	0.993
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parenthesis are clustered at the state level. $Nat\ Div_{s,t}$ is computed as the diversity among natives born in a different state than state s state where they reside. It is standardized to exhibit the same mean and standard deviation as $Mig\ Div_{s,t}$. $Nat\ Share_{s,t}$ is the immigration rate in state s where immigrants are natives born in a different state than s.

Two-stage Least-Squares

We now check whether our results hold when instrumenting our main variables of interest. Table 1.4 provides the results of our 2SLS regressions. In col. 1, 2, 6 and 7, we only instrument our main variable of interest, $MD^K_{s,t}$, and use the two IV strategies explained in subsection 1.3.2. The first one is a shift-share strategy, which uses the predicted diversity index based on the 1960 geographic structure of each diaspora. The second one is the gravity-like strategy *a la* Feyrer (2019). First-stage estimates are provided in Tables 1.A21 and 1.A22 in the Appendix. Then, in the remaining columns, we deal with the endogeneity of two other important regressors, the immigration rate ($m^K_{s,t}$) and the epidemiological term ($MY^K_{s,t}$). To do so, we use the gravity-like strategy *a la* Feyrer (2019) only.

Col. 1 and 2 in Table 1.4 confirm our previous findings for high-skilled diversity when $MD^H_{s,t}$ is instrumented only. The effect of $MD^H_{s,t}$ is always positive and highly significant. When using the shift-share strategy (col. 1), the magnitude of the coefficient is close to that of the FE regressions. The coefficient becomes larger under the gravity-like strategy *a*

la Feyrer (2019) (col. 2) even if it is not significantly different from the FE estimate. It is worth noticing that the instruments used in our IV regressions are valid. In particular, the Kleibergen-Paap F-stat of our second stage is always very large, and satisfies the Stock-Yogo critical values related to 10% maximal IV size. In addition, the F-test of the first stage is always above the critical value of 10. After instrumenting with the shift-share strategy, a 10% change in diversity induces a 5.1% change in GDP; equivalently, a one-standard-deviation change in high-skilled diversity increases GDP per capita by 2.8%, which is close to our benchmark results. As for low-skilled diversity, we find insignificant or weakly significant effects in col. 6 and 7. We conduct additional IV regressions to deal with the endogeneity of the immigration rate and of the epidemiological term in the remaining columns of Table 1.4. As the shift-share strategy does a poor job of predicting the immigration rate,[26] we only use the gravity-like strategy *a la* Feyrer (2019). Different combinations of endogenous regressors are considered and they all give rise to the same conclusions. In all specifications, the instrumental variables are strong. Our estimates for $MD^H_{s,t}$ are robust, and the magnitude of the coefficient is similar to the FE estimates. The effect of low-skilled diversity is always insignificant from col. 8 to 10. Under some specifications, we obtain a negative and significant epidemiological effect for both college-educated and low-skilled immigrants. Again, we find no evidence of a contamination effect. On the contrary, our epidemiological results are more in line with the effect of diversity: attracting immigrants from economically and culturally distant countries is beneficial for economic growth. Overall, our IV regressions support the view that increasing birthplace diversity among college-educated immigrants causes a rise in GDP per capita in the receiving state.

1.4.4 Channels of transmission

In this section, we investigate the channels of transmission through which birthplace diversity affects economic growth, as well as the conditions under which the effect of diversity is maximized. To mitigate endogeneity concerns, we use the gravity-like IV strategy *a la* Feyrer (2019) and instrument the diversity index as well as the share of immigrants.[27]

[26] The same problem arises in Alesina et al. (2016).

[27] Contrary to the shift-share, the gravity-like strategy provides valid instruments for the two variables of interest.

Table 1.4: 2SLS regressions under different IV strategies.
(Dep= $log(y_{s,t})$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Shift-Share	Feyrer	Feyrer	Feyrer	Feyrer	Shift-Share	Feyrer	Feyrer	Feyrer	Feyrer
Skill level	High skilled	High skilled	High skilled	High skilled	High skilled	Low skilled	Low skilled	Low skilled	Low skilled	Low skilled
$Mig\ Div_{s,t}$	0.511**	1.035***	0.853***	0.444**	0.726***	0.276*	0.186	0.089	0.105	0.096
	(0.213)	(0.308)	(0.278)	(0.200)	(0.259)	(0.157)	(0.132)	(0.142)	(0.089)	(0.151)
$Mig\ Share_{s,t}$	0.548*	0.878**	0.481	0.079	0.289	0.623*	0.527*	0.373	0.404	0.388
	(0.306)	(0.387)	(0.533)	(0.504)	(0.573)	(0.326)	(0.278)	(0.390)	(0.396)	(0.421)
$MY_{s,t}$			-0.122*	-0.208**	-0.215**			-0.106***	-0.098*	-0.099*
			(0.074)	(0.098)	(0.096)			(0.040)	(0.055)	(0.056)
$log(Pop_{s,t})$	-0.153**	-0.163**	-0.145*	-0.130*	-0.137*	-0.174**	-0.169**	-0.143*	-0.146*	-0.145*
	(0.074)	(0.073)	(0.076)	(0.078)	(0.078)	(0.080)	(0.076)	(0.076)	(0.083)	(0.075)
$log(Urb_{s,t})$	0.285**	0.283**	0.277**	0.276**	0.276**	0.378**	0.345**	0.307*	0.313*	0.309*
	(0.135)	(0.125)	(0.123)	(0.131)	(0.124)	(0.170)	(0.161)	(0.178)	(0.167)	(0.177)
$log(Hum_{s,t})$	0.772***	0.709***	0.705***	0.737***	0.702***	0.711***	0.776***	0.807***	0.801***	0.806***
	(0.190)	(0.199)	(0.193)	(0.186)	(0.192)	(0.210)	(0.218)	(0.211)	(0.191)	(0.210)
Endogenous regressors:										
$\widehat{Mig\ Div}_{s,t}$	✓	✓	✓		✓	✓	✓	✓		✓
$\widehat{MY}_{s,t}$				✓	✓				✓	✓
$\widehat{Mig\ Share}_{s,t}$			✓	✓	✓			✓	✓	✓
Observations	306	306	306	306	306	306	306	306	306	306
Nb. states	51	51	51	51	51	51	51	51	51	51
R-squared	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
K-P F-$Test^{\dagger}$	439.9	103.9	43.65	69.68	48.14	91.14	70.12	33.89	75.88	18.11
Stock Yogo	29.18/16.23	16.38/8.96	7.03/4.58	7.03/4.58	N.A	29.18/16.23	16.38/8.96	7.03/4.58	7.03/4.58	N.A

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. The specification is described in Eq. (1.3) and includes all fixed effects. We estimate it with 2SLS and rely on two IV strategies (the augmented shift-share and the gravity-like strategy *a la* Feyrer (2019)) to instrument the birthplace diversity index. Col. 1, 2, 3, 4 and 5 report estimates for high-skilled diversity; col. 6, 7, 8, 9 and 10 report estimates for low-skilled diversity. † Kleibergen-Paap F-statistic tests for weak identification (critical values from Stock-Yogo (2005) are given for 10%/15% maximal IV size). First-stage results are reported in Table 1.A21 and 1.A22 in the Appendix. Zero-stage regressions with instrumentation *a la* Feyrer (2019) are provided in Table 1.A20 in the Appendix. The sample includes the 50 US states and the District of Columbia from 1960 to 2010. The set of control variables includes the immigration rate ($Mig\ Share_{s,t}$), the log of population ($log(Pop_{s,t})$), the log of urbanization ($log(Urb_{s,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{s,t})$).

Alternative dependent variables

We first produce exploratory regressions using alternative dependent variables. These dependent variables include alternative outcomes of interest or intermediate variables influencing the level and/or growth rate of GDP per capita. We only focus on the effects of high-skilled diversity and of the share of immigrants in the college-educated population. We are aware that each of these alternative regressions should ideally include a specific set of appropriate controls. Nevertheless, for the sake of comparison, we use a common specification including the full set of fixed effects without time-varying controls. The results of these exploratory regressions are described in Table 1.5, which starts with the (benchmark) effect on GDP per capita in col. 1.

We first test whether diversity impacts the average employment rate in the receiving state. A positive effect on employment can be obtained if diversity stimulates linkages between sectors and firms, and/or if it increases labor productivity (Peri (2012)). We use census data from IPUMS to construct our aggregate measure of employment rate at the state level. The results in col. 2 show that high-skilled diversity does not significantly affect the log of the employment rate.

Secondly, we investigate the effect of high-skilled diversity on the performance of the R&D sector. A positive effect on innovation can be obtained if diversity enhances complementarities in R&D activities and improves the collective ability to solve problems. Such complementarities have been identified in the existing literature. Hunt and Gauthier-Loiselle (2010) show that the number of patents per capita increases with the number of college-educated immigrants in US states. Looking at the US Age of Mass Migration, Sequeira et al. (2017) find that immigration increases the number of patents registered by natives and foreigners. We use data from the US Patent and Trademark Office and compute the number of patents per capita for each state and for each decade. In col. 3, we find that high-skilled diversity increases the log number of patents per capita.

Thirdly, we examine whether diversity affects social capital. It can be the case that the economic gains from diversity are partly compensated for by a negative effect on trust, reciprocity or cooperation between people. We proxy social capital using data on violence and other criminal offences. We extract data from the US Department of Justice on the numbers of violent crimes and property crimes per capita, by state and by decade.[28] In col. 4 and 5, we show that high-skilled diversity does not significantly affect the crime rate.

[28]Violent crimes include murders, non negligent manslaughters, rapes, robberies and aggravated assaults. Property crimes include burglaries, larceny-thefts and motor vehicle thefts, among others.

Table 1.5: Transmissions channels for high-skilled diversity
IV gravity-like strategy *a la* Feyrer (2019) (Dep= $log(y_{s,t})$)

	(1) GDP cap (log)	(2) Employment (log)	(3) Patents (log)	(4) Violent crime (log)	(5) Property crime (log)
$Mig\ Div_{s,t}$	1.063***	0.056	3.011**	0.894	0.923*
	(0.377)	(0.066)	(1.461)	(1.279)	(0.531)
$Mig\ Share_{s,t}$	0.054	-0.131	3.277*	-1.531	-1.912**
	(0.755)	(0.081)	(1.963)	(1.696)	(0.893)
Observations	306	306	306	306	306
R-squared	0.991	0.773	0.315	0.822	0.848
Nb. states	51	51	51	51	51
Time fixed effects	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes
K-P F-$Test^{\dagger}$	49.70	49.70	49.70	49.70	49.70
Stock Yogo	7.03/4.58	7.03/4.58	7.03/4.58	7.03/4.58	7.03/4.58

	(6) Wages HS (log)*	(7) Wages LS (log)*	(8) Diversity occ. HS	(9) Diversity occ. LS
$Mig\ Div_{s,t}$	0.553*	0.718**	-0.004	0.001
	(0.286)	(0.347)	(0.013)	(0.040)
$Mig\ Share_{s,t}$	0.685*	0.555	0.014	-0.148***
	(0.359)	(0.364)	(0.014)	(0.054)
Observations	255	255	306	306
Nb. states	51	51	51	51
R-squared	0.824	0.608	0.911	0.384
Time fixed effects	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes
K-P F-$Test^{\dagger}$	30.02	30.02	49.70	49.70
Stock Yogo	7.03/4.58	7.03/4.58	7.03/4.58	7.03/4.58

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. The sample includes the 50 US states and the District of Columbia from 1960 to 2010. (*) When the logarithm of weekly wages is used as a dependent variable, the sample is reduced to 255 observations since data on the number of weeks worked are not included in the 1960 US census. Data for GDP are provided by the Bureau of Economic Analysis. Employment data are computed using IPUMS data and report the number of workers in the total labor force. Patents are the decennial total number of utility patents per capita (patents for invention) granted by the U.S. Patent and Trademark Office. Data on crime are taken from the U.S. Department of justice. Violent crime reports the total number of violent crimes per capita in a given state including murder and non negligent manslaughter, legacy and revised rape, robbery and aggravated assault. Property crime reports the total number of property crimes per capita in a given state including burglary, larceny-theft and motor vehicle theft. Following Borjas (2015), wages are measured as the log of natives' weekly wages (males and females aged 15 to 64). Diversity Occ. is the probability that two randomly drawn individuals in a state are working in two different occupations. † Kleibergen-Paap F-statistic tests for weak identification (critical values from Stock-Yogo (2005) are given for 10%/15% maximal IV size). We use the gravity-like strategy *a la* Feyrer (2019) to jointly instrument the birthplace diversity index and the share of immigrants.

Interestingly, the level of high-skilled immigration is negatively associated with the number of property crimes.

Fourthly, we assess whether the rise in GDP per capita benefits native workers or immigrants only. Recent studies on the labor market effect of immigration show that native and immigrant workers in a given skill-experience group are imperfect substitutes (Ottaviano

and Peri (2012), Manacorda et al. (2012), Card (2012)). The estimated elasticity of substitution varies across countries. We test whether birthplace diversity increases the level of labor market complementarities between immigrants and natives. To do so, we use the log of natives' earnings as a dependent variable. In line with Borjas (2015), we use census data to compute the average weekly wages of native workers by state, by education level, and for the years 1970 to 2010. Data on the number of weeks worked are not available for the census year 1960. Our sample includes all natives aged 15 to 64 irrespective of their gender. In col. 6 and 7, we find a positive and significant effect of high-skilled diversity on the log of natives' earnings. The effect is larger and more significant for low-skilled natives than for the highly skilled. Overall, these results suggest that diversity governs the size of skill complementarities between native and foreign workers.

Finally, to push the analysis of complementarities one step further, we investigate whether birthplace diversity increases the level of occupational diversity in the state of destination. Workers with a foreign background are more likely to start new businesses and to overcome labor shortages in certain sectors or occupations. Using IPUMS data, we compute an index of occupational fractionalization by state and by decade. This index measures the probability that two randomly drawn workers are employed in two different occupations. As many occupations are skill-specific, we compute separate indices for the two skill groups. In col. 8 and 9, we find that the effect of birthplace diversity on occupational fractionalization is insignificant. This result can be driven by the fact that occupational data are available at a low level of detail.

Drivers of skill complementarities

Our Herfindahl index does not account for the economic, cultural, linguistic, or institutional distance between the origin countries and the US. When adding separate epidemiological terms, we obtain insignificant or negative effects. This suggests that attracting migrants from distant countries is beneficial for growth on average. We now explore whether the interaction between diversity and origin country characteristics matters by using the Greenberg diversity indices defined in Eq. 1.5. The Greenberg index accounts for the cultural and economic distances between the origin country and the US (with weights equal to θ_1 and θ_2, respectively). When θ_1 increases (for a given θ_2), the Greenberg index over-weights immigrants from culturally distant countries (i.e. from genetically or linguistically distant countries). When θ_2 increases (for a given θ_1), it over-weights immigrants from richer countries (i.e. from less economically distant countries). Hence, we estimate our benchmark specification using the Greenberg (augmented) diversity index under various combinations

of θ_1 and θ_2 (ranging from -10 to +10). This helps us identify whether and how origin country characteristics influence the size of skill complementarities between immigrants and natives. For every combination of θ_1 and θ_2, we instrument the Greenberg index using the gravity-like strategy *a la* Feyrer (2019).[29]

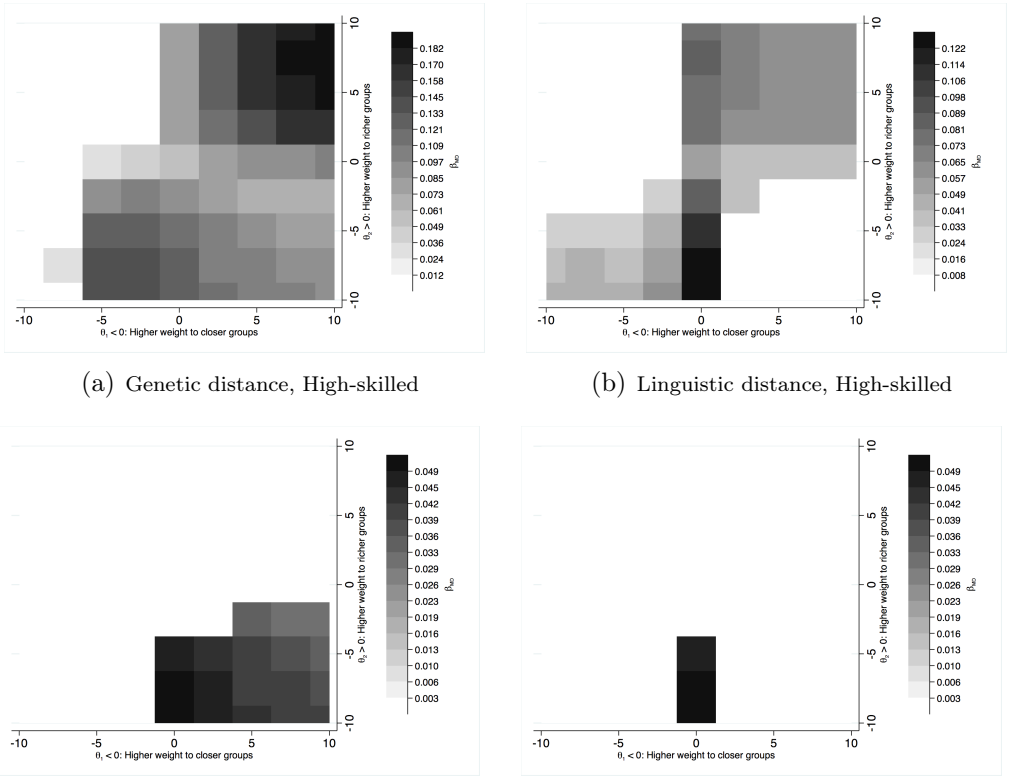
The results for high-skilled diversity are presented in Figures 1.5(a) and 1.5(b) (i.e. the top panel); the results for low-skilled diversity are presented in Figures 1.5(c) and 1.5(d) (i.e. the bottom panel). In the left panel, we use genetic distance as a proxy for cultural distance; in the right panel, we use linguistic distance. We only plot the standardized coefficients of augmented diversity that are statistically significant at the 5% level. The shade of grey is proportional to the magnitude of the effect: darker cells correspond to a greater effect of diversity (i.e., the combinations of θ_1 and θ_2 generating greater effects on income per capita). Conversely, lighter cells correspond to a smaller effect, and the coefficients that are not significant at the 5% level do not appear (white color).

Figures 1.5(a) and 1.5(b) show the results for high-skilled diversity. When holding cultural distance constant ($\theta_1 = 0$), over-weighting migrants from economically distant countries ($\theta_2 < 0$) increases the significance and the magnitude of the effect of diversity. This suggests that hosting immigrants from economically distant countries brings complementary skills to the state of destination. On the contrary, over-weighting immigrants from richer countries ($\theta_2 > 0$) leads to smaller results. In the same vein, when holding economic distance constant ($\theta_2 = 0$), over-weighting genetically or linguistically distant groups ($\theta_1 > 0$) increases the significance and magnitude of the effect of diversity. This suggests that cultural distance increases the effect of diversity. At $\theta_2 = 0$, overweighting migrants that are linguistically close to US citizens leads to insignificant results.

Economic and cultural distances are correlated. Nevertheless, looking at the interactions between them sheds light on the benefits from diversity and on the size of potential skill complementarities. Over-weighting groups that are richer (i.e., economically closer) and culturally closer to natives always induces insignificant results (top-left area of the figure). On the contrary, the greatest gains from diversity are obtained when considering groups that are economically close to and culturally distant from natives (top-right area of the figures). To a lesser extent, it is also growth-enhancing to attract high-skilled immigrants from culturally close and economically distant countries (bottom-left area of the figure). Hence, cultural distance is beneficial for the highly skilled when they originate from richer countries. Overall, in line with our skill complementarity hypothesis, diversity is beneficial when immigrants originate from economically OR culturally distant countries (but not both).

[29] In all regressions the Kleibergen-Paap F-statistic is above the critical values from Stock-Yogo (2005).

Figure 1.4: Augmented Diversity Index



(a) Genetic distance, High-skilled

(b) Linguistic distance, High-skilled

(c) Genetic distance, Low-skilled

(d) Linguistic distance, Low-skilled

Notes: Diversity among immigrants is defined as in Eq. (1.5). Each squared cell represents one estimated coefficient for $MD_{s,t}^{K,Aug}$, which itself corresponds to a particular combination of θ_1 and θ_2. Blank squared cells represent insignificant coefficients at the 5% level. Each shaded cell corresponds to a significant coefficient at the 5% level, and darker cells mean larger estimated coefficients. Data on genetic distance are obtained from Spolaore and Wacziarg (2009). Data on linguistic distance are obtained from Head et al. (2010). We use a dummy equal to one if a language is spoken by at least 9% of the population in both countries. Source: Authors' elaboration on IPUMS data.

These results hold when using genetic or linguistic distance as a cultural proxy.

These findings partially hold for low-skilled diversity. Figures 1.5(c) and 1.5(d) show that low-skilled diversity can also be growth-enhancing when immigrants come from economically distant countries. On the contrary, the effect is always insignificant when over-weighting immigrants from richer countries. Significant and positive effects are obtained when immigrants originate from poor and genetically distant countries ($\theta_2 < 0$ and $\theta_1 > 0$). This result is not confirmed when using linguistic distance. Contrary to high-skilled diversity, we do not find evidence of a positive effect of attracting low-skilled migrants from countries that are economically close but culturally similar. Cultural proximity among the low skilled is never

growth-enhancing, although we have not found any significant and negative effect associated with high cultural distance.

1.5 Conclusions

This paper empirically investigates the impact of birthplace diversity on economic growth. We use a rich data set that covers all US states in the post-World War II period. Compared to existing studies, we take advantage of the availability of panel data to conduct a large set of robustness checks and to better deal with unobserved heterogeneity and other endogeneity issues. We systematically test for skill-specific effects of diversity. Using a full set of fixed effects and combining various identification strategies, we show that diversity among college-educated immigrants positively and monotonically affects the macroeconomic performance of US states. This result is highly robust to the measurement of diversity, to sample selection, to specification choices and to the instrumentation strategy. Contrary to existing studies, diversity among less educated immigrants induces insignificant effects (or much smaller gains in certain limited circumstances). The latter result holds when we control for the share of undocumented migrants in this group. Still, we find no evidence of an immigration-driven contamination by bad economic and/or institutional conditions in origin countries.

Further important innovations are that we explore the channels of transmission through which diversity affects economic growth, and we identify the conditions under which the gains from high-skilled diversity are maximized. We provide converging evidence pointing at the existence of skill complementarities between workers trained in different countries. These synergies are observed on the labor market, and in the R&D sector. We show that high-skilled diversity increases the decennial number of patents per capita and the average weekly income of low-skilled and high-skilled natives. The effect is maximized when immigrants originate from economically or culturally distant countries (but not both), and when they acquired part of their secondary education abroad and their college education in the US. In terms of magnitude, a 10% increase in diversity among college-educated immigrants raises GDP per capita by about 6%, which is greater than in recent studies. This implies that high-skilled diversity explains 3.5% of the output rise between 1960 and 2010 in the US, and about 4% of the current output gap between the least and most diverse states.

Obviously, focusing on the US states over a specific period raises the problem of external validity of our results. For example, there is no guarantee that diversity shocks propelled by Africa-to-Europe migration or by intra-European mobility have generated the same economic

responses. However, the cross-country analysis conducted in the online appendix suggests that similar mechanisms are likely to operate in the other OECD countries.

1.6 Appendix

1.6.1 Descriptive statistics and data sources

We provide additional descriptive statistics as well as additional details on the data sources and definitions. Table 1.A1 provides the list of origin countries used in the paper. Table 1.A2 lists the variables used in the empirical analysis with their definition and source. Table 1.A3 provides the list of US states and their average characteristics. Figure 1.A1 depicts the immigration rates of various group of immigrants in the US. Figure 1.A2 and 1.A3 depict the trends in birthplace diversity at the state level. Finally, Table 1.A4 reports the Pearson correlations between alternative diversity measures.

Variable description and data sources

Table 1.A1: List of origin countries (195).

Afghanistan, Albania, Algeria, Andorra, Angola, Antigua-Barbuda, Argentina, Armenia, Australia, Austria, Azerbaijan, Bahamas, Bahrain, Bangladesh, Barbados, Belarus, Belgium, Belize, Benin, Bhutan, Bolivia, Bosnia, Botswana, Brazil, Brunei, Bulgaria, Burkina Faso, Burma (Myanmar), Burundi, Cambodia, Cameroon, Canada, Cape Verde, Central African Republic, Chad, Chile, China, China, Hong Kong SAR, China, Macao SAR, Colombia, Comoros, Congo, Dem. Rep. of the, Congo, Rep. of the, Costa Rica, Cote d'Ivoire, Croatia, Cuba, Cyprus, Czech Republic, Denmark, Djibouti, Dominica, Dominican Republic, East Timor, Ecuador, Egypt, El Salvador, Equatorial Guinea, Eritrea, Estonia, Ethiopia, Fiji, Finland, France, Gabon, Gambia, Georgia, Germany, Ghana, Greece, Grenada, Guatemala, Guinea, Guinea-Bissau, Guyana/British Guiana, Haiti, Holy See (Vatican City), Honduras, Hungary, Iceland, India, Indonesia, Iran, Iraq, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kazakhstan, Kenya, Kiribati, Korea, Kuwait, Kyrgyzstan, Laos, Latvia, Lebanon, Lesotho, Liberia, Libya, Liechtenstein, Lithuania, Luxembourg, Macedonia, Madagascar, Malawi, Malaysia, Maldives, Mali, Malta, Marshall Islands, Mauritania, Mauritius, Mexico, Micronesia, Moldova, Monaco, Mongolia, Morocco, Mozambique, Namibia, Nauru, Nepal, Netherlands, New Zealand, Nicaragua, Niger, Nigeria, Norway, Occupied Palestinian Territory, Oman, Pakistan, Palau, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Qatar, Romania, Russia, Rwanda, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Samoa, San Marino, Sao Tome and Principe, Saudi Arabia, Senegal, Serbia and Montenegro, Seychelles, Sierra Leone, Singapore, Slovakia, Slovenia, Solomon Islands, Somalia, South Africa, Spain, Sri Lanka, Sudan, Suriname, Swaziland, Sweden, Switzerland, Syria, Taiwan, Tajikistan, Tanzania, Thailand, Togo, Tonga, Trinidad and Tobago, Tunisia, Turkey, Turkmenistan, Tuvalu, Uganda, Ukraine, United Arab Emirates, United Kingdom, Unites States, Uruguay, Uzbekistan, Vanuatu, Venezuela, Vietnam, Yemen, Zambia, Zimbabwe.

Table 1.A2: Sources and definitions

Variable	Description	Definition	Source
$y_{s,t}$	Gross Domestic product	Gross domestic product (GDP) per capita.	Bureau of Economic Analysis.
$TD^K_{s,t}$	Birthplace diversity among residents	Probability that two randomly-drawn residents in state s have different countries of birth.	Authors' calculation on IPUMS-US data.
$MD^K_{s,t}$	Birthplace diversity among immigrants	Probability that two randomly-drawn immigrants in state s have different countries of birth.	Authors' calculation on IPUMPS-US data.
$MY^K_{s,t}$	Immigration-driven norm among immigrants	Weighted average outcome y in immigrant's origin countries where the weights are immigrants' share in the total immigrants' population in the destination region r.	Authors' calculation on IPUMS-US data.
$TP^K_{s,t}$	Polarization index among residents	Index that captures how the birthplace distribution in a population is far from the bimodal distribution.	Authors' calculation on IPUMS-US data.
$MP^K_{s,t}$	Polarization index among immigrants	Index that captures how the birthplace distribution in a immigrant's population is far from the bimodal distribution.	Authors' calculation on IPUMS-US data.
$k^K_{i,s,t}$	Share of immigrants	Number of individuals born in country i and living in state s as percentage of the total population of state s at year t.	IPUMS-US data.
$\widehat{k}^K_{i,s,t}$	Share of immigrants	Share of immigrants from origin country i in the total immigrant population of state s.	IPUMS-US data.
$m^K_{s,t}$	Immigration rate	Ratio of the total stock of foreign-born individuals to the total population of state s at year t.	Author's calculation on IPUMS-US.
$Hum_{s,t}$	Average education	Average education level.	Authors' calculation IPUMS data.
$Pop_{s,t}$	Population	Population of state s at year t.	IPUMS-US data.
$Urb_{s,t}$	Urbanization	Urban Percentage of the Population for States.	U.S. Census Bureau.
Gravity model:			
$Stock_{i,s,t}$	Stock of immigrants	Number of individuals born in country i and living in state s at year t.	IPUMS-US data.
$Distance_{i,s}$	Distance	Great-circle distance between the capital city of the origin countries i and the capital of the destination state s for US states.	Authors' calculation.
$Bord_{Canada,s}$	Common border	Dummy equal to 1 if Canada and State s share a common border and 0 otherwise.	Authors' elaboration.
$Bord_{Mexico,s}$	Common border	Dummy equal to 1 if Mexico and State s share a common border and 0 otherwise.	Authors' calculation.

Source: Authors' elaboration.

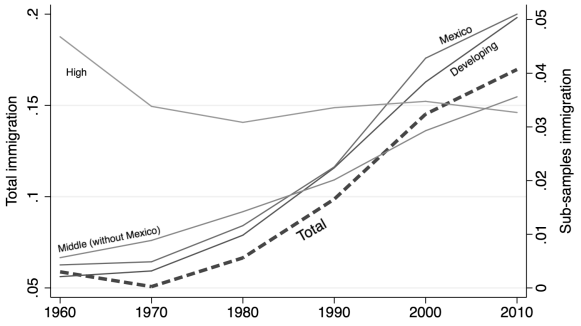
Table 1.A3: List of US states and descriptive statistics

States	$log(y_{s,t})$	$MD^{A}_{s,t}$	$MD^{H}_{s,t}$	$MD^{L}_{s,t}$	$m^{A}_{s,t}$	$log(Pop_{s,t})$	$log(Urb_{s,t})$	$log(Hum_{s,t})$
Alabama	9,256	0,900	0,923	0,846	0,021	14,715	4,060	1,722
Alaska	10,053	0,894	0,880	0,882	0,064	12,561	4,071	1,855
Arizona	9,484	0,690	0,906	0,554	0,107	14,440	4,428	1,806
Arkansas	9,209	0,871	0,914	0,811	0,024	14,154	3,930	1,706
California	9,731	0,845	0,941	0,734	0,221	16,618	4,519	1,839
Colorado	9,616	0,872	0,934	0,778	0,071	14,490	4,393	1,888
Connecticut	9,759	0,940	0,952	0,928	0,116	14,518	4,402	1,856
Delaware	9,847	0,930	0,932	0,889	0,055	12,919	4,303	1,827
District of Columbia	10,564	0,960	0,968	0,903	0,106	13,016	4,605	1,893
Florida	9,424	0,869	0,889	0,856	0,143	15,675	4,431	1,794
Georgia	9,458	0,909	0,939	0,854	0,052	15,174	4,164	1,741
Hawaii	9,657	0,745	0,828	0,711	0,172	13,347	4,457	1,846
Idaho	9,336	0,811	0,869	0,714	0,046	13,299	4,057	1,824
Illinois	9,675	0,883	0,947	0,792	0,109	15,807	4,438	1,816
Indiana	9,501	0,917	0,950	0,852	0,032	15,066	4,197	1,784
Iowa	9,476	0,909	0,945	0,856	0,029	14,379	4,077	1,826
Kansas	9,470	0,878	0,940	0,791	0,044	14,237	4,219	1,849
Kentucky	9,369	0,905	0,930	0,859	0,019	14,640	3,953	1,690
Louisiana	9,501	0,948	0,958	0,925	0,028	14,728	4,229	1,721
Maine	9,319	0,656	0,796	0,590	0,047	13,481	3,812	1,802
Maryland	9,592	0,955	0,960	0,938	0,088	14,872	4,389	1,836
Massachusetts	9,700	0,930	0,951	0,911	0,122	15,148	4,461	1,873
Michigan	9,561	0,927	0,928	0,910	0,060	15,568	4,290	1,809
Minnesota	9,580	0,938	0,950	0,907	0,046	14,789	4,222	1,855
Mississippi	9,119	0,917	0,898	0,879	0,015	14,242	3,820	1,701
Missouri	9,490	0,939	0,950	0,913	0,028	14,969	4,232	1,785
Montana	9,344	0,873	0,877	0,859	0,029	13,113	3,973	1,845
Nebraska	9,510	0,895	0,937	0,838	0,039	13,792	4,164	1,846
Nevada	9,759	0,863	0,923	0,802	0,125	13,342	4,439	1,818
New Hampshire	9,470	0,809	0,887	0,772	0,057	13,315	4,028	1,843
New Jersey	9,717	0,949	0,951	0,941	0,159	15,394	4,509	1,832
New Mexico	9,477	0,660	0,893	0,509	0,071	13,655	4,278	1,796
New York	9,805	0,954	0,964	0,944	0,189	16,276	4,453	1,827
North Carolina	9,475	0,905	0,948	0,845	0,039	15,232	3,929	1,737
North Dakota	9,362	0,873	0,834	0,857	0,029	12,886	3,889	1,817
Ohio	9,531	0,951	0,951	0,940	0,037	15,732	4,320	1,799
Oklahoma	9,360	0,883	0,942	0,796	0,036	14,445	4,193	1,798
Oregon	9,511	0,866	0,920	0,788	0,072	14,342	4,262	1,852
Pennsylvania	9,505	0,945	0,952	0,933	0,046	15,844	4,287	1,796
Rhode Island	9,478	0,888	0,931	0,858	0,109	13,346	4,477	1,790
South Carolina	9,273	0,903	0,924	0,857	0,028	14,548	3,981	1,716
South Dakota	9,326	0,911	0,901	0,893	0,021	12,971	3,868	1,809
Tennessee	9,375	0,916	0,943	0,863	0,025	14,920	4,099	1,720
Texas	9,569	0,610	0,917	0,452	0,113	16,091	4,385	1,764
Utah	9,450	0,883	0,925	0,822	0,066	13,739	4,432	1,880
Vermont	9,387	0,754	0,852	0,679	0,053	12,690	3,570	1,839
Virginia	9,560	0,953	0,952	0,940	0,070	15,132	4,201	1,802
Washington	9,709	0,897	0,912	0,853	0,098	14,862	4,331	1,870
West Virginia	9,248	0,933	0,915	0,910	0,013	13,958	3,700	1,700
Wisconsin	9,504	0,909	0,946	0,849	0,038	14,910	4,194	1,818
Wyoming	9,755	0,885	0,894	0,818	0,033	12,513	4,134	1,844

Note: Average from 1960 to 2010. Source: Authors' elaboration on IPUMS-US data.

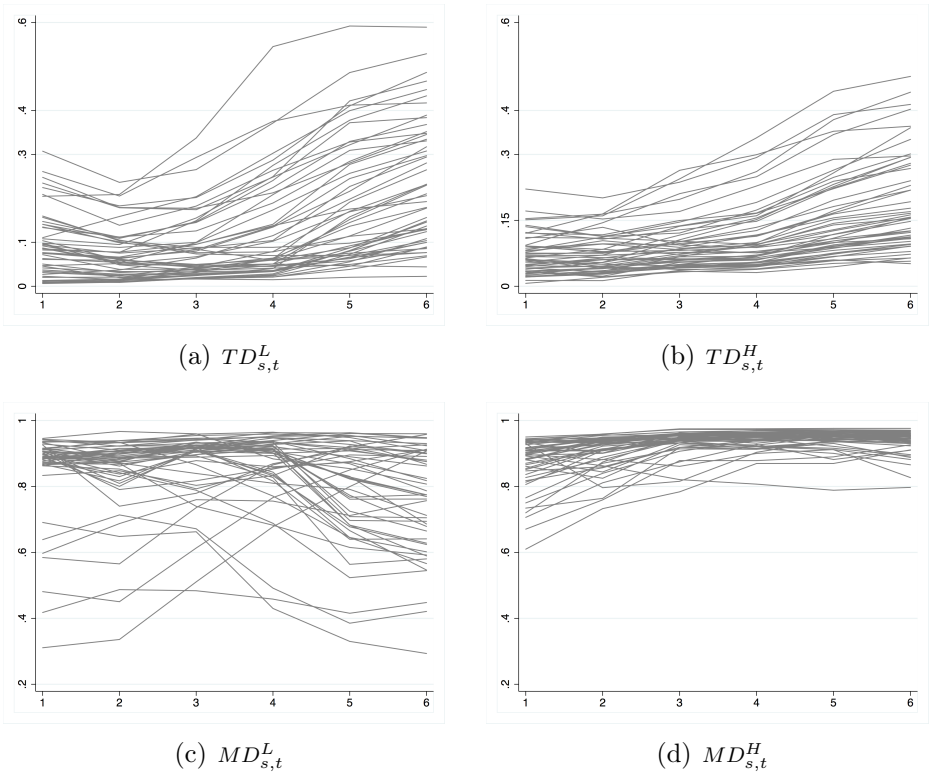
Trends in immigration and birthplace diversity in the US

Figure 1.A1: US immigration rate, 1960-2010 (as percentage of total population)



Notes: The "Total immigration rate" is defined as the ratio of the total stock of foreign-born individuals to the total population of the destination country or region. The "Immigration rate from high, developing and middle income countries" is defined as the ratio of the total stock of foreign-born individuals originating respectively from high, developing and middle countries to the total population of the destination country or region. The definition of a high, developing and middle income countries follows the World Bank classification of 2015. Source: Authors' elaboration on IPUMS-US data.

Figure 1.A2: Trends in birtplace diversity in the US states

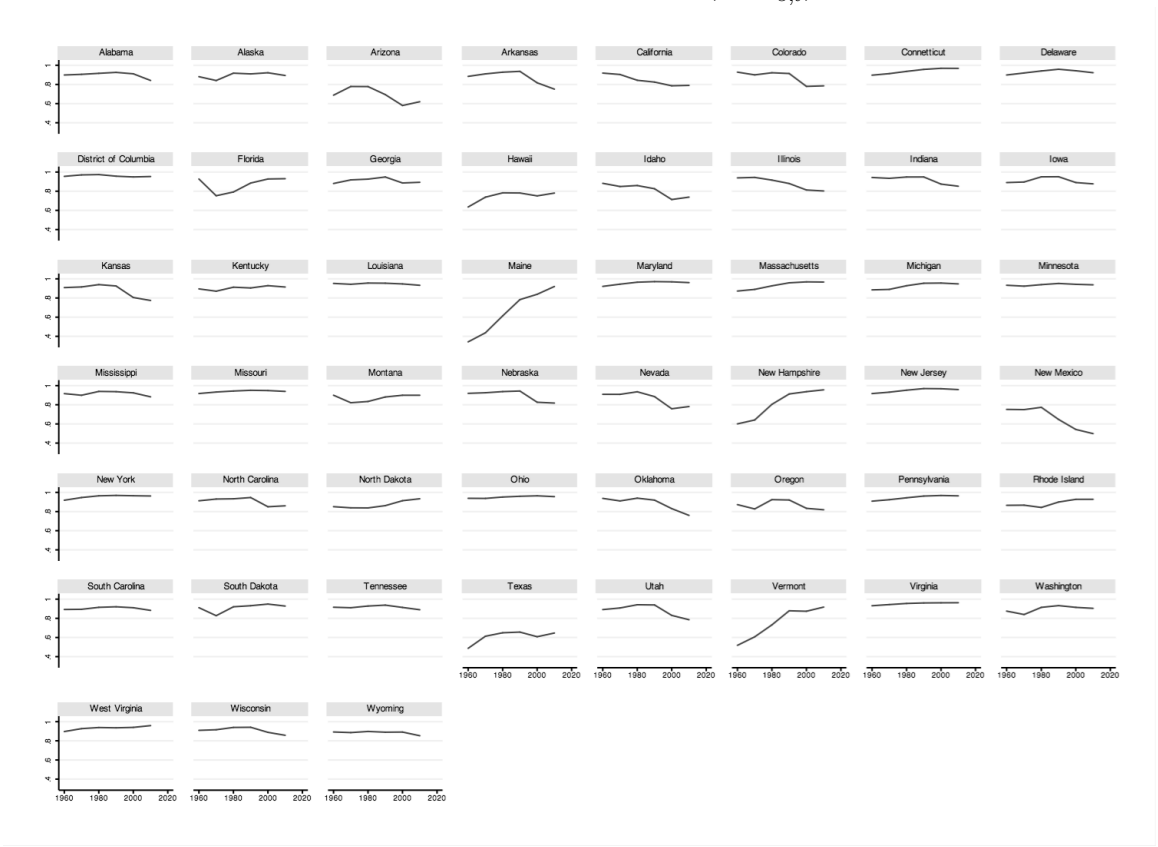


(a) $TD^L_{s,t}$ (b) $TD^H_{s,t}$

(c) $MD^L_{s,t}$ (d) $MD^H_{s,t}$

Notes: Diversity among residents is defined as in Eq. (1) in the paper. Diversity among immigrants is defined as in Eq (2) in the paper. Source: Authors' elaboration on IPUMS-US data.

Figure 1.A3: Diversity among immigrants ($MD^A_{s,t}$) in the US states



Notes: Diversity among residents is defined as in Eq. (1.1). Diversity among immigrants is defined as in Eq. (1.2) in the paper. Source: Authors' elaboration on IPUMS-US data.

Correlations between birthplace diversity indices

Table 1.A4: Pearson correlations between diversity measures

Variables	$TD^A_{s,t}$	$MD^A_{s,t}$	$TP^A_{s,t}$	$MP^A_{s,t}$	$MD^{A,G}_{s,t}$	$MY^A_{s,t}$	$m^A_{s,t}$
$TD^A_{s,t}$	1.000						
$MD^A_{s,t}$	-0.194***	1.000					
$TP^A_{s,t}$	0.987***	-0.244***	1.000				
$MP^A_{s,t}$	0.146***	-0.892***	0.190***	1.000			
$MD^{A,G}_{s,t}$	0.314***	0.393***	0.298***	-0.449***	1.000		
$MY^A_{s,t}$	-0.261***	-0.166***	-0.253***	0.189***	-0.182***	1.000	
$m^A_{s,t}$	0.998***	-0.196***	0.977***	0.152***	0.306***	-0.256**	1.000
Variables	$TD^H_{s,t}$	$MD^H_{s,t}$	$TP^H_{s,t}$	$MP^H_{s,t}$	$MD^{H,G}_{s,t}$	$MY^H_{s,t}$	$m^H_{s,t}$
$TD^H_{s,t}$	1.000						
$MD^H_{s,t}$	0.169***	1.000					
$TP^H_{s,t}$	0.990***	0.178***	1.000				
$MP^H_{s,t}$	-0.237***	-0.968***	-0.253***	1.000			
$MD^{H,G}_{s,t}$	0.430***	0.685***	0.439***	-0.733***	1.000		
$MY^H_{s,t}$	-0.128**	-0.189***	-0.116**	0.186***	-0.094	1.000	
$m^H_{s,t}$	0.999***	0.158***	0.984***	-0.224***	0.422***	-0.130**	1.000
Variables	$TD^L_{s,t}$	$MD^L_{s,t}$	$TP^L_{s,t}$	$MP^L_{s,t}$	$MD^{L,G}_{s,t}$	$MY^L_{s,t}$	$m^L_{s,t}$
$TD^L_{s,t}$	1.000						
$MD^L_{s,t}$	-0.340***	1.000					
$TP^L_{s,t}$	0.986***	-0.413***	1.000				
$MP^L_{s,t}$	0.238***	-0.828***	0.297***	1.000			
$MD^{L,G}_{s,t}$	0.194***	0.377***	0.156***	-0.364***	1.000		
$MY^L_{s,t}$	-0.322***	-0.148**	-0.313***	0.148**	-0.283***	1.000	
$m^L_{s,t}$	0.995***	-0.349***	0.972***	0.246***	0.178***	-0.303***	1.000

Notes: *** p<0.01, ** p<0.05. Source: Authors' elaboration on IPUMS data.

Other data Sources

The sources of our migration data were described in Section 1.2. In this subsection, we describe the data sources used to construct our dependent variables, the set of control variables, and the set of instruments. Table 1.A5 summarizes the descriptive statistics of our main variables. More details on our data sources and variable definitions are available in

Table 1.A2 in the Appendix. The GDP data ($y_{s,t}$) are provided by the Bureau of Economic Analysis for US states. The population data by age are taken from the IPUMS database. We consider the population aged 15 to 64 ($Pop_{s,t}$) in the regressions. The US Bureau of Census also provides the data on urbanization rates for US states ($Urb_{s,t}$); the urbanization rate measures the percentage of the population living in urbanized areas, and urban clusters are defined in terms of population size and density. As for human capital ($Hum_{s,t}$), we compute the average educational attainment of the working-age population using the IPUMS database.

As far as the set of instruments is concerned, the data on geographic distance between origin countries and US states are computed using the latitude and the longitude of the capital city of each US state and each country. Such data are available from the Infoplease website and the United Nations (2014) which have allowed us to compute a bilateral matrix of great-circle distances between US state capital cities and countries.[30]

Table 1.A5: Summary statistics 1960-2010

	Mean	s.d.	Min	Max
$TD^A_{s,t}$	0.126	0.105	0.006	0.548
$MD^A_{s,t}$	0.879	0.099	0.342	0.974
$m^A_{s,t}$	0.068	0.061	0.003	0.347
$TD^H_{s,t}$	0.116	0.087	0.007	0.478
$MD^H_{s,t}$	0.921	0.054	0.610	0.976
$m^H_{s,t}$	0.061	0.049	0.003	0.281
$TD^L_{s,t}$	0.134	0.121	0.006	0.592
$MD^L_{s,t}$	0.827	0.141	0.293	0.967
$m^L_{s,t}$	0.074	0.073	0.003	0.417
$log(y_{s,t})$	9.534	1.018	7.587	12.058
$log(Pop_{s,t})$	14.390	1.068	11.831	17.042
$log(Urb_{s,t})$	4.201	0.245	3.472	4.605
$log(Hum_{s,t})$	1.806	0.156	1.360	2.072

Source: Authors' elaboration on IPUMS-US data.

[30] See `http://www.infoplease.com/ipa/A0001796.html` (accessed on December, 2017).

Supplementary regression tables

We report the comprehensive tables of the regressions included in the paper. Table 1.A7 and 1.A8 report robustness checks of fixed effects regressions under alternative samples. Table 1.A9 replicates our benchmark specification at the Commuting Zone (CZ) level. As GDP per capita is not available at the CZ level, we use the log of the average weekly earnings as a dependent variable. Table 1.A10 and 1.A11 show that our results are stable when controlling for the share of the ten largest groups of immigrants. Table 1.A12 describes the sensitivity of our results to alternative diversity indices. Table 1.A13 replicates our analysis considering alternative education categories. Table 1.A14 investigates the role of undocumented migrants in governing the skill-specific effects of diversity. Table 1.A15 shows that our results are highly robust to the exclusion of control variables. Table 1.A16 investigates whether immigrants who arrived in the US at different ages generate different levels of skill complementarity with the native workforce. It includes alternative indices which alternatively exclude from the immigrant population the individuals who arrived in the US before a given age threshold, which ranges from 5 to 25 in one-year intervals. Table 1.A17 investigates the robustness of our results to region-by-census fixed effects which control for variations in regional business cycles. Table 1.A18 investigates which spatial decomposition is rich enough to capture the effects of diversity on economic growth; it compares diversity indices computed using shares by country, by continents, or by broad geographic regions of origin.

Table 1.A6: Robustness of FE regressions for low-skilled diversity (Dep= $log(y_{s,t})$)

	Sub-sample (1.A8)						CZ (1.A9)
	(1) Full-sample	(2) 1970-2000	(3) No Top5	(4) No Bot5	(5) No Mex	(6) No Q1	(7) CZ
$Mig\ Div_{s,t}$	0.141	0.130	0.228**	0.128	0.109	0.015	0.171**
	(0.086)	(0.097)	(0.093)	(0.092)	(0.091)	(0.115)	(0.207)
$Mig\ Share_{s,t}$	0.481*	0.691**	0.938**	0.448	0.474	0.647**	0.997***
	(0.282)	(0.272)	(0.418)	(0.288)	(0.392)	(0.276)	(0.350)
Observations	306	204	276	276	282	228	3,688
Nb. states/CZs	51	51	46	46	47	38	741
R-squared	0.993	0.989	0.993	0.993	0.993	0.993	0.891

	10 Largest (1.A11)	Quadratic (1.A12)	Educ. level (1.A13)			Legal status (1.A14)	
	(8)	(9)	(10) No school	(11) Primary	(12) Secondary	(13) Docum.	(14) Undoc.
$Mig\ Div_{s,t}$	0.276**	0.705**	0.038	0.033	0.120	0.043	-0.017
	(0.104)	(0.293)	(0.032)	(0.061)	(0.101)	(0.112)	(0.045)
$Mig\ Share_{s,t}$	0.058	0.504*	0.152	0.165	0.432	1.107**	3.482*
	(0.274)	(0.281)	(0.100)	(0.101)	(0.301)	(0.424)	(1.795)
$(Mig\ Div_{s,t})^2$		-0.391*					
		(0.218)					
Observations	306	306	306	306	306	204	204
Nb. states	51	51	51	51	51	51	51
R-squared	0.994	0.993	0.993	0.993	0.993	0.979	0.979

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state (or CZ) level. B-indexed numbers in parentheses refer to full Tables provided in the Appendix. All models (except col. 7) include the full vector of controls (not shown) with the log of population ($log(Pop_{s,t})$), the log of urbanization ($log(Urb_{s,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{s,t})$) as well as time and state fixed effects. * The sample for col. 7 includes the geographical area of the 50 US states and the District of Columbia from 1970 to 2010. The dependent variable is the logarithm of the average wage of white US natives between 40-50.

Table 1.A7: Robustness of FE regressions for high-skilled diversity
Alternative sub-samples (Dep= $log(y_{s,t})$)

	(1) Full Sample	(2) 1970-2000	(3) No Top5	(4) No Bot5	(5) No Mex	(6) No $Q1$
$MD^H_{s,t}$	0.616***	0.870***	0.725***	0.672***	0.630***	0.596**
	(0.160)	(0.321)	(0.174)	(0.170)	(0.170)	(0.288)
$m^H_{s,t}$	0.614*	1.140**	1.317**	0.613*	0.541	0.765**
	(0.315)	(0.459)	(0.529)	(0.323)	(0.397)	(0.365)
$log(Pop_{s,t})$	-0.155**	0.002	-0.187**	-0.160**	-0.158*	-0.182**
	(0.075)	(0.075)	(0.082)	(0.073)	(0.088)	(0.074)
$log(Urb_{s,t})$	0.285**	0.290	0.260*	0.300**	0.295**	0.198
	(0.135)	(0.187)	(0.140)	(0.147)	(0.138)	(0.151)
$log(Hum_{s,t})$	0.759***	1.251***	0.692***	0.945***	0.731***	0.797***
	(0.197)	(0.310)	(0.213)	(0.183)	(0.224)	(0.233)
Constant	7.348***	4.309***	7.870***	7.030***	7.373***	8.107***
	(1.262)	(1.584)	(1.387)	(1.250)	(1.398)	(1.313)
Observations	306	204	276	276	282	228
Nb. states	51	51	46	46	47	38
R-squared	0.993	0.990	0.993	0.993	0.993	0.995
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. The specification is described in Eq.(1.3) and includes all fixed effects. Col. 1 reports the results from Table 1.1. In col. 2, we exclude observations for the years 1960 and 2010. In col. 3 and 4, we exclude the five US states with the greatest or smallest immigration shares. In col. 5, we exclude US states located on the US-Mexican border. In col. 6, we exclude the lowest quartile in terms of immigrant population. The set of control variables includes the immigration rate ($m^K_{s,t}$), the log of population ($log(Pop_{s,t})$), the log of urbanization ($log(Urb_{s,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{s,t})$).

Table 1.A8: Robustness of FE regressions for low-skilled diversity
Alternative sub-samples (Dep= $log(y_{s,t})$)

	(1) Full Sample	(2) 1970-2000	(3) No Top5	(4) No Bot5	(5) No Mex	(6) No $Q1$
$MD^L_{s,t}$	0.141	0.130	0.228**	0.128	0.109	0.015
	(0.086)	(0.097)	(0.093)	(0.092)	(0.091)	(0.115)
$m^L_{s,t}$	0.481*	0.691**	0.938**	0.448	0.474	0.647**
	(0.282)	(0.272)	(0.418)	(0.288)	(0.392)	(0.276)
$log(Pop_{s,t})$	-0.166**	0.004	-0.200**	-0.169**	-0.157	-0.197**
	(0.081)	(0.086)	(0.090)	(0.080)	(0.095)	(0.076)
$log(Urb_{s,t})$	0.329**	0.323	0.315*	0.341*	0.313*	0.360**
	(0.163)	(0.200)	(0.165)	(0.182)	(0.168)	(0.166)
$log(Hum_{s,t})$	0.807***	1.309***	0.743***	0.964***	0.785***	0.919***
	(0.205)	(0.343)	(0.212)	(0.204)	(0.215)	(0.264)
Constant	7.662***	4.738***	8.183***	7.439***	7.660***	7.971***
	(1.263)	(1.514)	(1.370)	(1.259)	(1.429)	(1.158)
Observations	306	204	276	276	282	228
Nb. states	51	51	46	46	47	38
R-squared	0.993	0.989	0.993	0.993	0.993	0.996
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. The specification is described in Eq.(1.3) and includes all fixed effects. Col. 1 reports the results from Table 1.1. In col. 2, we exclude observations for the years 1960 and 2010. In col. 3 and 4, we exclude the five US states with the greatest or smallest immigration shares. In col. 5, we exclude US states located on the US-Mexican border. In col. 6, we exclude the lowest quartile in terms of immigration rate. The set of control variables includes the immigration rate ($m^K_{s,t}$), the log of population ($log(Pop_{s,t})$), the log of urbanization ($log(Urb_{s,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{s,t})$).

Table 1.A9: Robustness of FE and IV regressions to spatial scale
Results by skill group at the CZ level (Dep= $log(Wage^K_{CZ,t})$)

	(1) Fixed-effects $K = H$	(2) Fixed-effects $K = L$	(3) Shift-Share $K = H$	(4) Shift-Share $K = L$
$MD^K_{s,t}$	0.319**	0.171**	0.372***	-0.317
	(0.150)	(0.076)	(0.128)	(0.207)
$m^K_{s,t}$	1.907***	0.997***	1.913***	0.439
	(0.567)	(0.350)	(0.561)	(0.438)
Constant	5.395***	5.528***		
	(0.110)	(0.063)		
Observations	3,688	3,688	3,688	3,688
R-squared	0.895	0.894	0.895	0.891
Number of CZ	741	741	741	741
Time fixed effects	Yes	Yes	Yes	Yes
CZ fixed effects	Yes	Yes	Yes	Yes
F-test (IV)			2305	501.3

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the Commuting zone (CZ) level. The specification includes all fixed effects. The sample includes the 50 US states and the District of Columbia from 1970 to 2010. The dependent variable = $log(Wage^K_{CZs,t})$ is logarithm of the average wage of white US natives between 40-50 which is not affected by discrimination, following Ottaviano and Peri (2006). Commuting zones are computed following Dorn (2009).

Table 1.A10: Robustness of FE regressions for high-skilled diversity
Ten largest US immigrants group in 2010 (Dep= $log(y_{s,t})$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
i	Mexico	India	Philippines	China	Vietnam	El Salvador	Cuba	Korea	Dominican Rep.	Guatemala	10 largest
$MD^H_{s,t}$	0.649***	0.616***	0.602***	0.577***	0.615***	0.613***	0.627***	0.617***	0.607***	0.621***	0.617***
	(0.176)	(0.161)	(0.164)	(0.149)	(0.160)	(0.156)	(0.158)	(0.160)	(0.156)	(0.163)	(0.169)
$m^H_{s,t}$	0.673**	0.743**	0.626*	0.638**	0.615*	0.500	0.631**	0.628*	0.550	0.649**	0.726*
	(0.316)	(0.349)	(0.315)	(0.311)	(0.315)	(0.303)	(0.301)	(0.336)	(0.338)	(0.322)	(0.365)
$\widehat{k}^H_{i,s,t}$	0.214	-0.628*	-0.126	0.403	-0.035	2.033	0.063	-0.158	0.354	-0.813	
	(0.260)	(0.352)	(0.199)	(0.280)	(0.458)	(1.483)	(0.237)	(0.425)	(0.712)	(1.313)	
$log(Pop_{s,t})$	-0.166**	-0.174**	-0.150**	-0.144*	-0.155**	-0.157**	-0.155**	-0.156**	-0.151*	-0.156**	-0.169**
	(0.078)	(0.075)	(0.074)	(0.079)	(0.076)	(0.074)	(0.075)	(0.076)	(0.079)	(0.075)	(0.083)
$log(Urb_{s,t})$	0.282**	0.278**	0.283**	0.292**	0.286**	0.271*	0.285**	0.289**	0.296**	0.287**	0.282*
	(0.136)	(0.139)	(0.139)	(0.134)	(0.138)	(0.139)	(0.136)	(0.136)	(0.145)	(0.138)	(0.162)
$log(Hum_{s,t})$	0.807***	0.852***	0.743***	0.766***	0.758***	0.760***	0.766***	0.753***	0.752***	0.759***	0.895***
	(0.218)	(0.196)	(0.198)	(0.198)	(0.205)	(0.196)	(0.202)	(0.199)	(0.198)	(0.195)	(0.233)
Constant	7.404***	7.506***	7.328***	7.173***	7.349***	7.442***	7.327***	7.352***	7.263***	7.356***	7.341***
	(1.263)	(1.212)	(1.248)	(1.338)	(1.266)	(1.240)	(1.272)	(1.266)	(1.354)	(1.262)	(1.360)
Observations	306	306	306	306	306	306	306	306	306	306	306
Nb. states	51	51	51	51	51	51	51	51	51	51	51
R-squared	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. $\widehat{k}^H_{i,s,t}$ is the share of high-skilled immigrants from origin country i in the total immigrant population of state r. The sample includes the 50 US states and the District of Columbia from 1960 to 2010. Col. 11 includes $\widehat{k}^H_{i,s,t}$ for the 10 largest immigrant groups in the US. Coefficients for all the countries are not reported in col. 11 for space limitations.

Table 1.A11: Robustness of FE regressions for low-skilled diversity
Ten largest US immigrants group in 2010 (Dep= $log(y_{s,t})$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
i	Mexico	India	Philippines	China	Vietnam	El Salvador	Cuba	Korea	Dominican Rep.	Guatemala	10 largest
$MD^L_{s,t}$	0.294**	0.149*	0.159	0.213**	0.127	0.133	0.131	0.148	0.147*	0.142*	0.276**
	(0.111)	(0.086)	(0.097)	(0.093)	(0.093)	(0.082)	(0.089)	(0.090)	(0.086)	(0.085)	(0.104)
$m^L_{s,t}$	0.652**	0.454*	0.252	0.483*	0.484*	0.211	0.446	0.482*	0.500*	0.443	0.058
	(0.290)	(0.266)	(0.266)	(0.260)	(0.282)	(0.249)	(0.283)	(0.281)	(0.293)	(0.291)	(0.274)
$\widehat{k}^L_{i,s,t}$	0.173*	-1.781	-1.472*	-1.394*	0.250	0.620***	-0.232	-0.103	-0.087	0.351	
	(0.089)	(1.159)	(0.803)	(0.831)	(0.364)	(0.209)	(0.209)	(0.295)	(0.414)	(0.399)	
$log(Pop_{s,t})$	-0.187**	-0.172**	-0.151**	-0.190**	-0.161*	-0.127	-0.164**	-0.166**	-0.169*	-0.162*	-0.129*
	(0.078)	(0.079)	(0.071)	(0.075)	(0.084)	(0.077)	(0.080)	(0.081)	(0.086)	(0.082)	(0.069)
$log(Urb_{s,t})$	0.335*	0.307*	0.308**	0.317**	0.312*	0.320*	0.325**	0.336**	0.328*	0.329**	0.256*
	(0.167)	(0.163)	(0.147)	(0.153)	(0.174)	(0.162)	(0.161)	(0.153)	(0.164)	(0.162)	(0.149)
$log(Hum_{s,t})$	0.792***	0.839***	0.681***	0.810***	0.804***	0.733***	0.808***	0.801***	0.808***	0.777***	0.622***
	(0.203)	(0.209)	(0.217)	(0.207)	(0.202)	(0.211)	(0.200)	(0.204)	(0.206)	(0.207)	(0.218)
Constant	7.804***	7.786***	7.768***	7.999***	7.683***	7.292***	7.655***	7.637***	7.698***	7.659***	7.665***
	(1.273)	(1.229)	(1.166)	(1.172)	(1.267)	(1.097)	(1.254)	(1.262)	(1.326)	(1.261)	(0.975)
Observations	306	306	306	306	306	306	306	306	306	306	306
Nb. states	51	51	51	51	51	51	51	51	51	51	51
R-squared	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.994
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. $\widehat{k}^L_{i,s,t}$ is the share of low-skilled immigrants from origin country i in the total immigrant population of state r. The sample includes the 50 US states and the District of Columbia from 1960 to 2010. Col. 11 includes $\widehat{k}^L_{i,s,t}$ for the 10 largest immigrant groups in the US. Coefficients for all the countries are not reported in col. 11 for space limitations.

Table 1.A12: Robustness of FE estimates to alternative specifications and measures
Results by skill group (Dep= $log(y_{s,t})$)

	Quadratic		Polarization		Country Count				Without Mexico		Without Hispanic		Theil Index					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	$K=H$	$K=L$	$K=H$	$K=L$	$K=H$	$K=H$	$K=L$	$K=L$	$K=H$	$K=L$	$K=H$	$K=L$	$K=H$	$K=H$	$K=H$	$K=L$	$K=L$	$K=L$
$MD^K_{s,t}$	-0.131	0.705**				0.593***		0.133	0.641***	0.377**	0.718***	0.356*						
	(1.954)	(0.293)				(0.154)		(0.088)	(0.170)	(0.186)	(0.167)	(0.193)						
$(MD^K_{s,t})^2$	0.453	-0.391*																
	(1.202)	(0.218)																
$MP^K_{s,t}$			-0.291***	-0.025														
			(0.090)	(0.072)														
$CC^K_{s,t}$					-0.001	-0.001	-0.001	-0.001										
					(0.001)	(0.001)	(0.001)	(0.001)										
$m^{K,Mexican}_{s,t}$									-0.876	-0.598								
									(1.750)	(0.533)								
$m^{K,Hispanic}_{s,t}$											-1.639	-1.084						
											(1.231)	(0.825)						
$T^K_{s,t}$													-0.086***			-0.023		
													(0.027)			(0.022)		
$T^{B,K}_{s,t}$														-0.019			-0.011	
														(0.021)			(0.034)	
$T^{W,K}_{s,t}$															-0.065**			-0.018
															(0.031)			(0.018)
$m^K_{s,t}$	0.622*	0.504*	0.596*	0.352	0.211	0.590*	0.359	0.493	0.711**	0.885**	1.009***	1.251*	0.587*	0.318	0.190	0.431	0.341	0.393
	(0.314)	(0.281)	(0.306)	(0.271)	(0.347)	(0.336)	(0.277)	(0.281)	(0.345)	(0.426)	(0.374)	(0.639)	(0.319)	(0.349)	(0.334)	(0.276)	(0.287)	(0.267)
Observations			306	306	306	306	306	306	306	306	306	306	306	306	306	306	306	306
Nb. states			51	51	51	51	51	51	51	51	51	51	51	51	51	51	51	51
R-squared			0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993						
Controls			Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects			Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects			Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. Col. 1 and 2 tests for a quadratic specification in birthplace diversity with $MD^H_{s,t}$ and $(MD^H_{s,t})^2$. In col. 3 and 4, we replace birthplace diversity by a polarization index ($MP^H_{s,t}$). The sample includes the 50 US states and the District of Columbia from 1960 to 2010. The set of control variables includes the immigration rate ($m^K_{s,t}$), the log of population ($log(Pop_{s,t})$), the log of urbanization ($log(Urb_{s,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{s,t})$).

Table 1.A13: Robustness of FE regressions to alternative educational levels
(Dep= $log(y_{s,t})$)

	(1) $K = H$	(2) $S = Ph.D$	(3) $S = Tertiary$	(4) $K = L$	(5) $S = No\ school$	(6) $S = Primary$	(7) $S = Secondary$
$MD^K_{s,t}$	0.616***	0.262**	0.369***	0.141	0.038	0.033	0.120
	(0.160)	(0.103)	(0.136)	(0.086)	(0.032)	(0.061)	(0.101)
$m^K_{s,t}$	0.614*	0.256	0.372	0.481*	0.152	0.165	0.432
	(0.315)	(0.266)	(0.298)	(0.282)	(0.100)	(0.101)	(0.301)
$log(Pop_{s,t})$	-0.155**	-0.158**	-0.141*	-0.166**	-0.167**	-0.160**	-0.149*
	(0.075)	(0.077)	(0.077)	(0.081)	(0.076)	(0.076)	(0.080)
$log(Urb_{s,t})$	0.285**	0.287*	0.253*	0.329**	0.207	0.287	0.311*
	(0.135)	(0.145)	(0.148)	(0.163)	(0.168)	(0.173)	(0.163)
$log(Hum_{s,t})$	0.759***	0.763***	0.779***	0.807***	0.843***	1.004***	0.831***
	(0.197)	(0.205)	(0.206)	(0.205)	(0.210)	(0.206)	(0.209)
Constant	7.348***	7.702***	7.489***	7.662***	8.205***	7.552***	7.497***
	(1.262)	(1.244)	(1.290)	(1.263)	(1.342)	(1.197)	(1.300)
Observations	306	306	306	306	306	306	306
Nb. states	51	51	51	51	51	51	51
R-squared	0.993	0.993	0.993	0.993	0.993	0.993	0.993
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. Col 1 and 4 report our benchmark specifications from Table 2.

Table 1.A14: Robustness of FE regressions
Results by legal status and skill group (Dep= $log(y_{s,t})$)

	(1) $K=H$ All	(2) $K=H$ Legal	(3) $K=H$ Undoc.	(4) $K=L$ All	(5) $K=L$ Legal	(6) $K=L$ Undoc.
$MD^K_{s,t}$	0.813*	1.009**	-0.153	0.038	0.043	-0.017
	(0.438)	(0.473)	(0.127)	(0.100)	(0.112)	(0.045)
$m^K_{s,t}$	0.842*	0.959*	4.426**	0.957**	1.107**	3.482*
	(0.431)	(0.535)	(2.140)	(0.377)	(0.424)	(1.795)
$ln(Population_{s,t})$	-0.029	-0.028	0.013	-0.074	-0.057	-0.088
	(0.092)	(0.094)	(0.088)	(0.105)	(0.099)	(0.112)
$ln(Urban_{s,t})$	0.129	0.129	0.050	0.093	0.090	0.079
	(0.164)	(0.166)	(0.172)	(0.167)	(0.166)	(0.180)
$ln(College_{s,t})$	2.043***	1.997***	2.296***	2.508***	2.418***	2.671***
	(0.633)	(0.622)	(0.611)	(0.690)	(0.676)	(0.724)
Constant	4.762***	4.650**	4.839***	5.450***	5.366***	5.497***
	(1.723)	(1.758)	(1.647)	(1.630)	(1.634)	(1.706)
Observations	204	204	204	204	204	204
Nb. states	51	51	51	51	51	51
R-squared	0.979	0.979	0.979	0.979	0.979	0.979
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses adjusted for clustering at the state level. Col. 1 and 4 report the coefficient of the benchmark sample over the 1980-2010 period and for high-skilled and low-skilled immigrants, respectively. In col. 2 and 5, the diversity indices are computed for the legal immigrant population only. In col. 3 and 6, we use the undocumented immigrant population only.

Table 1.A15: Robustness of Pooled OLS, FE and IV regressions without controls
Results by skill group (Dep= $log(y_{s,t})$)

	(1) OLS $K=H$	(2) Fixed-effects $K=H$	(3) Shift-Share $K=H$	(4) Feyrer $K=H$	(5) OLS $K=L$	(6) Fixed-effects $K=L$	(7) Shift-Share $K=L$	(8) Feyrer $K=L$
$MD^K_{s,t}$	7.595***	0.664***	0.418*	1.079***	0.343	0.147**	0.009	0.130
	(0.909)	(0.209)	(0.234)	(0.376)	(0.498)	(0.063)	(0.120)	(0.192)
$m^K_{s,t}$	9.912***	-0.019	-0.168	0.231	6.963***	-0.123	-0.290	-0.143
	(1.421)	(0.452)	(0.452)	(0.529)	(1.053)	(0.314)	(0.334)	(0.468)
Observations	306	306	306	306	306	306	306	306
Nb. states	51	51	51	51	51	51	51	51
R-squared	0.452	0.992	0.992	0.991	0.236	0.991	0.991	0.991
Time fixed effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
States fixed effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
K-P F-$Test^{\dagger}$			383.45	107.45			100.82	47.31
Stock Yogo			16.38/ 8.96	16.38/ 8.96			16.38/ 8.96	16.38/ 8.96

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. The specification is described in Eq. (1.3). The sample includes the 50 US states and the District of Columbia from 1960 to 2010. We estimate 2SLS relying on two IV strategies (the augmented shift-share and the gravity-like strategy *a la* Feyrer (2019)) to instrument the birthplace diversity index. †Kleinbergenn-Paap F-statistic tests for weak identification (critical values from Stock-Yogo (2005) are given for 10%/15% maximal IV size).

Table 1.A16: Robustness of FE regressions to age of entry, 1970-2010

	$K=H$	$K=H$	$K=H$	$K=H$	$K=H$	$K=H$	$K=H$	$K=H$	$K=H$	$K=H$	$K=H$	$K=H$	$K=H$	$K=H$	$K=H$	$K=H$	$K=H$	$K=H$	$K=H$	$K=H$	$K=H$
Age of entry >	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
$MD^H_{s,t}$	0.733**	0.739**	0.703**	0.676**	0.694**	0.655**	0.645**	0.638**	0.564**	0.550**	0.548**	0.535**	0.545**	0.526**	0.414**	0.218***	0.210***	0.226**	0.244**	0.231**	0.216**
	(0.308)	(0.310)	(0.307)	(0.305)	(0.326)	(0.322)	(0.307)	(0.308)	(0.222)	(0.225)	(0.222)	(0.218)	(0.222)	(0.215)	(0.160)	(0.079)	(0.078)	(0.090)	(0.103)	(0.094)	(0.087)
$m^H_{s,t}$	0.633	0.641	0.647	0.652	0.677	0.659	0.681	0.690	0.699	0.695	0.714	0.736	0.735	0.771	0.752	0.660	0.665	0.682	0.745	0.762	0.765
	(0.385)	(0.395)	(0.408)	(0.414)	(0.433)	(0.433)	(0.445)	(0.449)	(0.441)	(0.444)	(0.454)	(0.462)	(0.470)	(0.494)	(0.498)	(0.487)	(0.522)	(0.547)	(0.591)	(0.649)	(0.733)
$ln(Population_{s,t})$	-0.087	-0.090	-0.090	-0.092	-0.094	-0.091	-0.092	-0.093	-0.086	-0.085	-0.085	-0.085	-0.085	-0.086	-0.080	-0.068	-0.067	-0.066	-0.066	-0.066	-0.065
	(0.066)	(0.067)	(0.067)	(0.067)	(0.067)	(0.067)	(0.068)	(0.068)	(0.067)	(0.067)	(0.067)	(0.067)	(0.067)	(0.068)	(0.068)	(0.068)	(0.068)	(0.068)	(0.067)	(0.067)	(0.067)
$ln(Urban_{s,t})$	0.226	0.229	0.218	0.225	0.216	0.218	0.217	0.219	0.204	0.205	0.215	0.221	0.213	0.211	0.207	0.190	0.196	0.190	0.183	0.184	0.181
	(0.154)	(0.154)	(0.152)	(0.154)	(0.152)	(0.154)	(0.150)	(0.151)	(0.148)	(0.149)	(0.149)	(0.150)	(0.149)	(0.148)	(0.149)	(0.157)	(0.157)	(0.157)	(0.155)	(0.158)	(0.158)
$ln(College_{s,t})$	0.887***	0.881***	0.906***	0.910***	0.915***	0.921***	0.924***	0.924***	0.951***	0.956***	0.948***	0.940***	0.946***	0.958***	0.970***	1.001***	0.996***	0.997***	1.000***	0.999***	1.001***
	(0.278)	(0.278)	(0.283)	(0.283)	(0.286)	(0.286)	(0.288)	(0.288)	(0.289)	(0.289)	(0.289)	(0.288)	(0.289)	(0.293)	(0.292)	(0.291)	(0.289)	(0.292)	(0.296)	(0.295)	(0.297)
Constant	6.589***	6.618***	6.661***	6.673***	6.721***	6.702***	6.724***	6.726***	6.712***	6.704***	6.682***	6.681***	6.691***	6.706***	6.732***	6.749***	6.726***	6.727***	6.743***	6.751***	6.756***
	(1.315)	(1.313)	(1.295)	(1.294)	(1.281)	(1.288)	(1.271)	(1.269)	(1.272)	(1.276)	(1.280)	(1.283)	(1.280)	(1.268)	(1.270)	(1.280)	(1.284)	(1.284)	(1.275)	(1.283)	(1.285)
R-squared	0.991	0.991	0.991	0.991	0.991	0.991	0.991	0.991	0.991	0.991	0.991	0.991	0.991	0.991	0.991	0.990	0.990	0.990	0.991	0.990	0.990
	$K=L$	$K=L$	$K=L$	$K=L$	$K=L$	$K=L$	$K=L$	$K=L$	$K=L$	$K=L$	$K=L$	$K=L$	$K=L$	$K=L$	$K=L$	$K=L$	$K=L$	$K=L$	$K=L$	$K=L$	$K=L$
Age of entry >	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
$MD^L_{s,t}$	0.062	0.059	0.060	0.057	0.044	0.043	0.039	0.037	0.037	0.029	0.027	0.029	0.022	0.018	0.020	0.027	0.027	0.022	0.022	0.030	0.026
	(0.071)	(0.071)	(0.072)	(0.071)	(0.073)	(0.072)	(0.071)	(0.072)	(0.071)	(0.073)	(0.073)	(0.075)	(0.077)	(0.077)	(0.077)	(0.077)	(0.078)	(0.078)	(0.075)	(0.080)	(0.082)
$m^L_{s,t}$	0.462*	0.459*	0.457*	0.458*	0.443	0.444	0.437	0.441	0.452	0.444	0.439	0.457	0.470	0.482	0.520	0.534	0.570	0.601	0.637	0.683	0.733
	(0.258)	(0.258)	(0.259)	(0.262)	(0.268)	(0.270)	(0.272)	(0.275)	(0.280)	(0.284)	(0.288)	(0.300)	(0.313)	(0.330)	(0.353)	(0.379)	(0.409)	(0.443)	(0.479)	(0.518)	(0.570)
$ln(Population_{s,t})$	-0.085	-0.085	-0.084	-0.083	-0.083	-0.082	-0.081	-0.081	-0.081	-0.080	-0.079	-0.079	-0.079	-0.078	-0.078	-0.077	-0.077	-0.076	-0.076	-0.076	-0.075
	(0.075)	(0.075)	(0.075)	(0.075)	(0.075)	(0.075)	(0.075)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)
$ln(Urban_{s,t})$	0.242	0.242	0.241	0.240	0.233	0.233	0.232	0.232	0.232	0.229	0.228	0.228	0.226	0.225	0.226	0.228	0.229	0.228	0.229	0.230	0.231
	(0.174)	(0.174)	(0.174)	(0.174)	(0.175)	(0.175)	(0.175)	(0.175)	(0.175)	(0.176)	(0.176)	(0.176)	(0.177)	(0.177)	(0.176)	(0.176)	(0.178)	(0.178)	(0.176)	(0.175)	(0.175)
$ln(College_{s,t})$	1.038***	1.037***	1.036***	1.038***	1.049***	1.047***	1.049***	1.052***	1.053***	1.057***	1.055***	1.051***	1.054***	1.051***	1.045***	1.027***	1.023***	1.025***	1.019***	1.010***	1.010***
	(0.307)	(0.308)	(0.310)	(0.309)	(0.312)	(0.311)	(0.311)	(0.312)	(0.311)	(0.313)	(0.312)	(0.312)	(0.313)	(0.311)	(0.310)	(0.307)	(0.306)	(0.307)	(0.306)	(0.308)	(0.309)
Constant	6.858***	6.857***	6.850***	6.848***	6.857***	6.853***	6.848***	6.847***	6.841***	6.843***	6.840***	6.841***	6.850***	6.851***	6.849***	6.847***	6.850***	6.852***	6.851***	6.854***	6.845***
	(1.278)	(1.277)	(1.277)	(1.278)	(1.280)	(1.281)	(1.283)	(1.283)	(1.283)	(1.284)	(1.287)	(1.288)	(1.290)	(1.292)	(1.291)	(1.294)	(1.294)	(1.295)	(1.295)	(1.295)	(1.293)
R-squared	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990
Observations	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255
Nb. states	51	51	51	51	51	51	51	51	51	51	51	51	51	51	51	51	51	51	51	51	51
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Source: Author's calculations on IPUMS-US data. Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses adjusted for clustering at the state level.

Table 1.A17: Regressions with region-by-year fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)
	East vs. West		4 Regions		9 Divisions	
	$K = H$	$S = L$	$K = H$	$K = L$	$K = H$	$S = L$
$MD^K_{x,t}$	0.567***	0.138	0.490***	0.054	0.320**	-0.086
	(0.170)	(0.086)	(0.164)	(0.099)	(0.148)	(0.122)
$m^K_{x,t}$	0.619**	0.528*	0.274	0.294	0.413	0.219
	(0.305)	(0.272)	(0.285)	(0.263)	(0.329)	(0.296)
$log(Pop_{x,t})$	-0.181**	-0.197**	-0.160	-0.177	-0.201**	-0.201**
	(0.073)	(0.077)	(0.100)	(0.106)	(0.092)	(0.099)
$log(Urb_{x,t})$	0.255*	0.291*	0.369**	0.411**	0.305*	0.298
	(0.145)	(0.167)	(0.163)	(0.179)	(0.179)	(0.213)
$log(Hum_{x,t})$	0.905***	0.992***	0.208	0.323	0.321	0.438
	(0.319)	(0.320)	(0.307)	(0.301)	(0.296)	(0.324)
Observations	306	306	306	306	306	306
Nb. of states	51	51	51	51	51	51
R-squared	0.994	0.994	0.994	0.994	0.995	0.995
Region × Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors adjusted for clustering at the states level. States are grouped as follow. East: Alaska, Arizona, California, Colorado, Hawaii, Idaho, Kansas, Montana, Nebraska, Nevada, New Mexico, North Dakota, Oklahoma, Oregon, South Dakota, Texas, Utah, Washington, Wyoming. West: Alabama, Arkansas, Connecticut, Delaware, District of Columbia, Florida, Georgia, Illinois, Indiana, Iowa, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, New Hampshire, New Jersey, New York, North Carolina, Ohio, Pennsylvania, Rhode Island, South Carolina, Tennessee, Vermont, Virginia, West Virginia, Wisconsin. Source: Author's calculations on IPUMS-US data.

Table 1.A18: FE regressions with aggregated group of migrants.
Results by skill group (Dep= $log(y_{x,t})$)

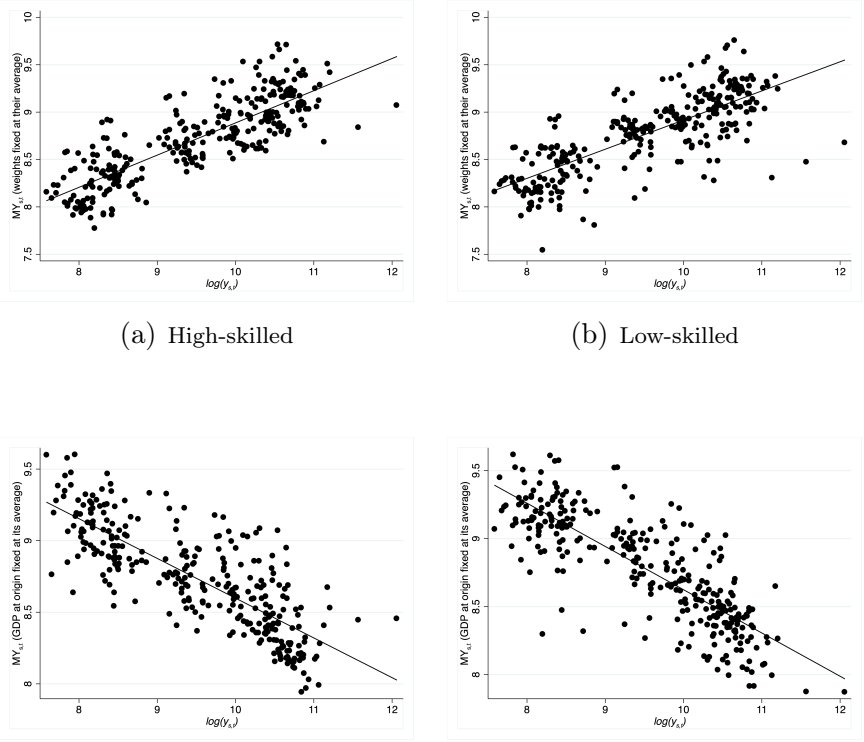
	(1) $K=H$ Continents	(2) $K=L$ Continents	(3) $K=H$ Regions	(4) $K=L$ Regions	(5) $K=H$ Country	(6) $K=L$ Country
$MD^K_{x,t}$	0.146	-0.079	0.305	0.099	0.616***	0.141
	(0.115)	(0.090)	(0.203)	(0.081)	(0.160)	(0.086)
$m^K_{x,t}$	0.295	0.236	0.472	0.474*	0.614*	0.481*
	(0.347)	(0.265)	(0.340)	(0.275)	(0.315)	(0.282)
$log(Pop_{x,t})$	-0.133	-0.169**	-0.144*	-0.159*	-0.155**	-0.166**
	(0.081)	(0.084)	(0.078)	(0.082)	(0.075)	(0.081)
$log(Urb_{x,t})$	0.290*	0.268	0.293*	0.313*	0.285**	0.329**
	(0.163)	(0.172)	(0.154)	(0.167)	(0.135)	(0.163)
$log(Hum_{x,t})$	0.819***	0.919***	0.777***	0.868***	0.759***	0.807***
	(0.209)	(0.202)	(0.204)	(0.202)	(0.197)	(0.205)
Constant	7.393***	7.952***	7.438***	7.584***	7.348***	7.662***
	(1.327)	(1.360)	(1.317)	(1.325)	(1.262)	(1.263)
Observations	306	306	306	306	306	306
Nb. states	51	51	51	51	51	51
R-squared	0.993	0.993	0.993	0.993	0.993	0.993
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. The sample includes the 50 US states and the District of Columbia from 1960 to 2010. Continent includes Africa, Americas, Asia, Europe and Oceania. Regions include Australia and New Zealand, Caribbean, Central America, Central Asia, Eastern Africa, Eastern Asia, Eastern Europe, Melanesia, Micronesia, Middle Africa, Northern Africa, Northern America, Northern Europe, Polynesia, South America, South-Eastern Asia, Southern Africa, Southern Asia, Southern Europe, Western Africa, Western Asia and Western Europe.

1.6.2 Epidemiological effects

We conduct additional robustness checks for the epidemiological effect. Figure 1.A4 depicts correlation between the epidemiological term and GDP per capita, keeping either GDP at origin or immigration rates as constant (at their 1960-2010 average level). Table 1.A19 reports results for additional specifications of the epidemiological term. We first compute $MY_{s,t}^{K}$ by keeping the immigration shares ($\widehat{k}_{i,s,t}^{K}$) constant, at their 1960-2010 average levels. Then, we keep the levels of GDP per capita at origin ($log(y_{i,t})$) constant, at their 1960-2010 average level. Finally, we combine annual data on GDP per capita at origin with individual data on the year of arrival in the US; each immigration share is multiplied by the average level of GDP per capita prevailing in the year of immigration to the US. The latter specification allows us to capture the norms and values that immigrants bring with them when they migrate. Due to data limitations, this variable cannot be computed for the year 1960.

Figure 1.A4: Cross-state correlation between epidemiological terms and GDP



(a) High-skilled (b) Low-skilled

(c) High-skilled (d) Low-skilled

Source: Authors' elaboration on IPUMS data.

Table 1.A19: Robustness of FE estimates to alternative definitions of the epidemiological term

Results by skill group (Dep= $log(y_{s,t})$)

	(1) $K=H$	(2) $K=H$	(3) $K=H$	(4) $K=H$	(5) $K=L$	(6) $K=L$	(7) $K=L$	(8) $K=L$
$MD^K_{s,t}$	0.531***	0.618***	0.538***	0.725***	0.104	0.141	0.101	0.006
	(0.159)	(0.162)	(0.157)	(0.249)	(0.085)	(0.085)	(0.086)	(0.085)
$MY^K_{s,t}$	-0.133*				-0.104**			
	(0.069)				(0.042)			
$MY^K_{s,t}$ (Const. $\widehat{k}^K_{i,s,t}$)		-0.120				0.001		
		(0.255)				(0.234)		
$MY^K_{s,t}$ (Const. $y_{i,t}$)			-0.110				-0.092**	
			(0.075)				(0.044)	
$MY^K_{s,t}$ ($y_{i,Entry}$)				-0.111***				-0.146**
				(0.039)				(0.057)
$m^K_{s,t}$	0.388	0.582*	0.441	0.539	0.412	0.481*	0.443	0.530**
	(0.366)	(0.300)	(0.368)	(0.328)	(0.283)	(0.273)	(0.285)	(0.241)
$log(Pop_{s,t})$	-0.144*	-0.157**	-0.143*	-0.080	-0.146*	-0.166**	-0.151*	-0.071
	(0.080)	(0.077)	(0.082)	(0.065)	(0.082)	(0.078)	(0.084)	(0.075)
$log(Urb_{s,t})$	0.282**	0.295**	0.273**	0.156	0.312*	0.329**	0.317*	0.194
	(0.135)	(0.130)	(0.134)	(0.138)	(0.173)	(0.163)	(0.172)	(0.169)
$log(Hum_{s,t})$	0.744***	0.740***	0.762***	1.007***	0.802***	0.807***	0.801***	1.108***
	(0.189)	(0.215)	(0.192)	(0.299)	(0.196)	(0.203)	(0.199)	(0.281)
Constant	8.455***	8.341***	8.299***	7.492***	8.379***	7.651***	8.384***	8.032***
	(1.229)	(3.006)	(1.164)	(1.273)	(1.317)	(2.708)	(1.316)	(1.320)
Observations	306	306	306	255	306	306	306	255
Nb. states	51	51	51	51	51	51	51	51
R-squared	0.993	0.993	0.993	0.991	0.993	0.993	0.993	0.991
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. The specification is described in Eq. (1.3) and includes all fixed effects. We supplement our benchmark specifications reported in col. 1 and 5 we alternatives definition of the epidemiological term ($MY^K_{s,t}$). We compute $MY^K_{s,t}$ (Const. $\widehat{k}^K_{i,s,t}$) by keeping the immigration shares constant, at their 1960-2010 average levels. We $MY^K_{s,t}$ (Const. $y_{i,t}$) by keeping the levels of GDP per capita at origin ($log(y_{i,t})$) constant, at their 1960-2010 average level. We compute $MY^K_{s,t}$ ($y_{i,Entry}$) combining annual data on GDP per capita at origin with individual data on the year of arrival in the US. Each immigration share is multiplied by the average level of GDP per capita prevailing in the year of immigration to the US which allows us to capture the norms and values that immigrants bring with them when they migrate. Due to data limitations, this variable cannot be computed for the year 1960. The sample includes the 50 US states and the District of Columbia from 1960 to 2010. The set of control variables includes the immigration rate ($m^K_{s,t}$), the log of population ($log(Pop_{s,t})$), the log of urbanization ($log(Urb_{s,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{s,t})$).

1.6.3 IV: zero- and first-stage regressions

We report the results of the zero-stage, pseudo-gravity regression *a la* Feyrer (2019) in Table 1.A20. In addition, Tables 1.A21 and 1.A21 present the first-stage regressions associated with the results described in section 1.4.3 of the paper (for high-skilled and low-skilled diversity, respectively).

Table 1.A20: Zero-stage estimates (PPML): gravity model *a la* Feyrer (2019)

	(1) $S = A$ $log(Stock_{i,s,t})$	(2) $K = H$ $log(Stock_{i,s,t})$	(3) $K = L$ $log(Stock_{i,s,t})$
$log(Dist_{i,s}) \times I_{1960}$	-1.666***	-1.480***	-1.741***
	(0.32)	(0.31)	(0.34)
$log(Dist_{i,s}) \times I_{1970}$	-1.786***	-1.463***	-1.954***
	(0.33)	(0.33)	(0.35)
$log(Dist_{i,s}) \times I_{1980}$	-1.760***	-1.349***	-2.033***
	(0.35)	(0.35)	(0.36)
$log(Dist_{i,s}) \times I_{1990}$	-1.733***	-1.280***	-2.112***
	(0.35)	(0.35)	(0.36)
$log(Dist_{i,s}) \times I_{2000}$	-1.821***	-1.241***	-2.250***
	(0.35)	(0.35)	(0.36)
$log(Dist_{i,s}) \times I_{2010}$	-1.827***	-1.276***	-2.309***
	(0.35)	(0.35)	(0.36)
$Bord_{Canada,s}$	3.605***	2.710***	4.352***
	(0.84)	(0.70)	(0.95)
$Bord_{Mexico,s}$	1.153***	0.999***	1.240***
	(0.20)	(0.24)	(0.21)
Constant	18.808***	15.445***	19.162***
	(2.96)	(2.89)	(3.13)
Observations	59364	59364	59364
R-squared	0.884	0.777	0.907
Year dummies	Yes	Yes	Yes
Origin dummies	Yes	Yes	Yes
Destination dummies	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Col.1 shows the results of the stocks of all migrants. Columns 2 and 3 illustrate the results for the college and low educated immigrants respectively. Standard errors in parentheses adjusted for clustering at the state/country-pair level. Source: Authors' elaboration on IPUMS data.

Table 1.A21: First-Stage regressions High-skilled (Alternative IV Strategies)

	(1) $MD^H_{s,t}$	(2) $MD^H_{s,t}$	(3) $MD^H_{s,t}$	(3) $m^H_{s,t}$	(4) $MY^H_{s,t}$	(4) $m^H_{s,t}$	(5) $MD^H_{s,t}$	(5) $MY^H_{s,t}$	(5) $m^H_{s,t}$
$MD^H_{s,t}$(Shift-share)	0.751***								
	(0.036)								
$MD^H_{s,t}$(Feyrer)		0.304***	0.302***	-0.003			0.318***	-0.236*	-0.002
		(0.030)	(0.032)	(0.013)			(0.027)	(0.093)	(0.014)
$m^H_{s,t}$ (Feyrer)			-0.921***	1.600***	-0.189	1.483***	-0.678***	0.513	1.528***
			(0.188)	(0.141)	(0.660)	(0.141)	(0.155)	(0.709)	(0.143)
$MY^H_{s,t}$ (Feyrer)					0.349***	-0.013***	0.031***	0.323***	-0.014***
					(0.020)	(0.004)	(0.008)	(0.020)	(0.004)
$MD^H_{s,t}$					-0.919***	-0.043			
					(0.237)	(0.027)			
$m^H_{s,t}$	-0.266	-0.524***							
	(0.137)	(0.108)							
$MY^H_{s,t}$			0.007	-0.020*					
			(0.021)	(0.009)					
$log(Pop_{s,t})$	-0.004	0.009	0.020	-0.016	0.002	-0.012	0.010	-0.011	-0.013
	(0.012)	(0.015)	(0.017)	(0.010)	(0.035)	(0.009)	(0.016)	(0.039)	(0.010)
$log(Urb_{s,t})$	0.005	-0.027	-0.007	-0.035*	0.058	-0.035*	-0.008	0.059	-0.036*
	(0.022)	(0.078)	(0.081)	(0.017)	(0.068)	(0.014)	(0.075)	(0.071)	(0.015)
$log(Hum_{s,t})$	0.067	0.081	0.148	-0.119	-0.044	-0.103	0.130	-0.175	-0.111
	(0.047)	(0.072)	(0.084)	(0.062)	(0.171)	(0.055)	(0.075)	(0.204)	(0.058)
Observations	306	306	306	306	306	306	306	306	306
R-squared	0.717	0.660	0.642	0.853	0.748	0.862	0.680	0.719	0.860
Angrist-Pischke F-test	439.93	103.94	87.26	99.19	233.71	108.93	138.18	237.69	95.81

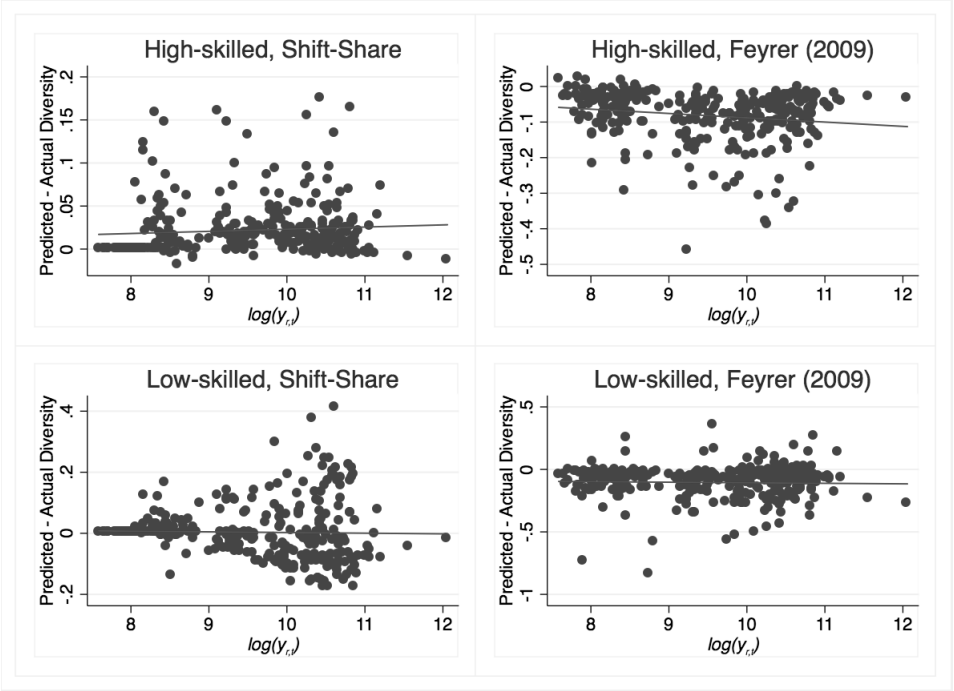
Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses. Source: Authors' elaboration on IPUMS-US data. This table provides the results of the first-stage regressions under different IV (Shift share and Feyrer (2019)).

Table 1.A22: First-Stage regressions Low-skilled (Alternative IV Strategies)

	(1) $MD^L_{s,t}$	(2) $MD^L_{s,t}$	(3) $MD^L_{s,t}$	(3) $m^L_{s,t}$	(4) $MY^L_{s,t}$	(4) $m^L_{s,t}$	(5) $MD^L_{s,t}$	(5) $MY^L_{s,t}$	(5) $m^L_{s,t}$
$MD^L_{s,t}$ (Shift-Share)	0.788***								
	(0.083)								
$MD^L_{s,t}$ (Feyrer)		0.416***	0.390***	0.022			0.393***	-0.177**	0.014
		(0.050)	(0.048)	(0.016)			(0.047)	(0.058)	(0.013)
$m^L_{s,t}$ (Feyrer)			-1.166***	1.379***	-0.592	1.337***	-1.150***	-0.608	1.346***
			(0.304)	(0.129)	(0.428)	(0.120)	(0.300)	(0.446)	(0.119)
$MY^L_{s,t}$ (Feyrer)					0.385***	-0.006	0.008	0.371***	-0.005
					(0.027)	(0.004)	(0.013)	(0.027)	(0.004)
$MD^L_{s,t}$					-0.213	0.014			
					(0.109)	(0.022)			
$m^L_{s,t}$	-0.863***	-0.603*							
	(0.229)	(0.232)							
$MY^L_{s,t}$			0.006	0.010					
			(0.039)	(0.014)					
$log(Pop_{s,t})$	-0.024	-0.052	-0.028	-0.008	0.180**	0.000	-0.029	0.219***	-0.003
	(0.048)	(0.040)	(0.041)	(0.013)	(0.059)	(0.010)	(0.039)	(0.061)	(0.012)
$log(Urb_{s,t})$	-0.043	-0.206	-0.164	-0.049*	-0.151	-0.050*	-0.166	-0.137	-0.050*
	(0.091)	(0.141)	(0.138)	(0.021)	(0.111)	(0.022)	(0.138)	(0.100)	(0.022)
$log(Hum_{s,t})$	0.639**	0.232	0.311	-0.145	-0.114	-0.133*	0.306	-0.070	-0.139
	(0.192)	(0.215)	(0.213)	(0.079)	(0.272)	(0.064)	(0.212)	(0.310)	(0.077)
Observations	306	306	306	306	306	306	306	306	306
R-squared	0.519	0.604	0.617	0.881	0.762	0.880	0.618	0.769	0.881
Angrist-Pischke F-test	91.14	70.12	73.14	75.92	198.22	120.62	76.08	183.86	85.90

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses. Source: Authors' elaboration on IPUMS-US data. This table provides the results of the first-stage regressions under different IV (Shift share and Feyrer).

Figure 1.A5: Difference between predicted and actual diversity indices



Source: Authors' elaboration on IPUMS-US data.

Using common regressors to predict high-skilled and low-skilled immigrant stocks, one might be concerned that our gravity-like strategy does not allow to identify skill-specific effects. This is not the case since the correlation between the predicted skill-specific diversity indices only equals 0.29 (once year and state fixed effects are netted out). This is due to the fact that the changes in moving costs between 1960 and 2010 induced heterogeneous effects on low-skilled and high-skilled migrants. Note that the correlation between predicted skill-specific indices equals 0.49 when using the shift-share strategy. Heterogeneity is driven here by the difference in the spatial structures of low-skilled and high-skilled immigrant stocks in 1960. In both cases, there is enough variation to identify skill-specific effects of birthplace diversity on economic growth. Furthermore, Figure 1.A5 shows that gap between observed and predicted diversity levels is not always positively correlated with the economic performance of the US state. In contrast with standard cross-country analyses, some fast-growing states are likely to attract more immigrants from less distant countries (e.g., Mexico). In other words, at the state level, it is a priori unclear whether reverse causation leads to an upward or downward bias in OLS estimation.

1.6.4 Self-selection of immigrants

Two methods are used to check whether our results are driven by the self-selection of immigrants. First, we follow Alesina et al. (2016) and assume that self-selection on unobserved productive traits can be proxied by self-selection on observable characteristics. We construct an index of selection on educational attainment in Eq. (1.12), and we include it in the benchmark specification. For each origin/state dyad, we divide the share of college graduates in the migration stock by the share of college graduates in the population of the source country. Then, for each US state, we compute the weighted average of these dyadic self-selection ratios, using the origin-specific shares of immigrants as weights. The selection index is written:

$$Selection_{s,t} = \sum_{i=1}^{I} \left[\frac{(Stock^{H}_{i,s,t})/(Stock_{i,s,t})}{(Pop^{H}_{i,t})/(Pop_{i,t})} \times \widehat{k}_{i,s,t} \right] \tag{1.12}$$

where $Stock^{H}_{i,s,t}/Stock_{i,s,t}$ is the share of college graduates in the dyadic stock of immigrants from origin country i to state s at time t, $Pop^{H}_{i,t}/Pop_{i,t}$ is the share of college graduates in the population of the origin country provided by Barro and Lee (2013), and $\widehat{k}_{i,s,t}$ is the share of immigrants originating from country i in state s at time t.

The mean of our selection index equals 11.2 (the min equals 2.7, the max equals 43.7, and the standard deviation equals 5.4). This confirms that migrants positively self-select along education levels. This result is in line with Docquier et al. (2007), Grogger and Hanson (2011) and Belot and Hatton (2012). In Table 1.A23(a) we include our self-selection index in col. 2 and 5. We also include its interaction with birthplace diversity in col. 3 and 6. Controlling for self-selection on educational attainment does not invalidate our results. In col. 2 and 5, the effect of high-skilled diversity remains positive and highly significant, and the effect of low-skilled diversity is small and less significant. The interaction term in col. 3 is not significant and adding it hardly affects the significance of the diversity effects for the high skilled. In the same way, while the interaction term is significant for the low skilled in col. 6, the total effect of diversity is always small and only significant when considering the first decile of the self-selection index.[31]

Still, it could be possible that self-selection on unobserved traits is poorly captured by self-selection on educational attainment. Thus, we extend the analysis above and follow the methodology proposed by Oster (2019). When the R^2 is close to 1 (i.e. when unexplained

[31]It can also be argued that migrant networks reduce the self-selection of recent immigrants, as evidenced by McKenzie and Rapoport (2010). Nevertheless, Tables 1.A10 and 1.A11 in the appendix show that our results are preserved when accounting for the size of the 10 largest migrant diasporas in the US.

Table 1.A23: Accounting for selection on observables and unobservables (Dep=$log(y_{s,t})$)

Table 1.A23(a): Selection on observables

	(1)	(2)	(3)	(4)	(5)	(6)
Skill level	High-Skilled	High-Skilled	High-Skilled	Low-Skilled	Low-Skilled	Low-Skilled
$Mig\ Div_{s,t}$	0.616***	0.713***	0.850***	0.141	0.170*	0.384***
	(0.160)	(0.186)	(0.254)	(0.086)	(0.101)	(0.116)
$Mig\ Share_{s,t}$	0.614*	0.691**	0.666**	0.481*	0.478*	0.477*
	(0.315)	(0.340)	(0.330)	(0.282)	(0.271)	(0.266)
$Selection_{s,t}$		-0.003	0.019		-0.002	0.033***
		(0.003)	(0.026)		(0.003)	(0.011)
$Mig\ Div_{s,t} \times Selection_{s,t}$			-0.023			-0.038***
			(0.029)			(0.012)
$log(Pop_{s,t})$	-0.155**	-0.160**	-0.157**	-0.166**	-0.167**	-0.145*
	(0.075)	(0.073)	(0.072)	(0.081)	(0.079)	(0.080)
$log(Urb_{s,t})$	0.285**	0.263*	0.281*	0.329**	0.326**	0.361**
	(0.135)	(0.134)	(0.142)	(0.163)	(0.162)	(0.147)
$log(Hum_{s,t})$	0.759***	0.721***	0.729***	0.807***	0.762***	0.753***
	(0.197)	(0.206)	(0.202)	(0.205)	(0.221)	(0.210)
Constant	7.348***	7.512***	7.253***	7.662***	7.749***	7.112***
	(1.262)	(1.234)	(1.249)	(1.263)	(1.258)	(1.228)
Observations	306	306	306	306	306	306
Nb. states	51	51	51	51	51	51
R-squared	0.993	0.993	0.993	0.993	0.993	0.993
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Table 1.A23(b): Selection on unobservables, Oster (2019)

	(7)	(8)	(9)	(10)	(11)	(12)
	No controls	FEs	FEs & $Mig\ Share^{K}_{s,t}$	FEs & Controls	δ for $\beta = 0$	*Id. set*
	$(s.d.)[R^2]$	$(s.d.)[R^2]$	$(s.d.)[R^2]$	$(s.d.)[R^2]$	& $R_{max} = 1$	given $R_{max} = 1$
$Mig\ Div_{s,t}$	9.02***	0.668***	0.663***	0.616***	2.71	[0.395, 0.616]
	(1.365)[0.229]	(0.202)[0.992]	(0.208)[0.992]	(0.160)[0.993]		
$Mig\ Div_{s,t}$	-0.92	0.168**	0.146**	0.141	-4.16	[0.107, 0.141]
	(0.612)[0.012]	(0.075)[0.991]	(0.063)[0.992]	(0.085)[0.993]		

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. Table 1.A23(a): The set of control variables includes the immigration rate ($Mig\ Share_{s,t}$), the log of population ($log(Pop_{s,t})$), the log of urbanization ($log(Urb_{s,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{s,t})$). Table 1.A23(b): col. 7 to 10 report the effect of skill specific diversity with their associated Std. Error and R^2. Col. 10 shows the results when we introduce the log of population ($log(Pop_{s,t})$), the log of urbanization ($log(Urb_{s,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{s,t})$) as additional covariates. Col. 11 shows the value of δ which produces $\beta = 0$ given the value of $R_{max} = 1$. The identified set in col. 12 is bounded by $\hat{\beta}$ when $\delta = 0$ (no bias-adjustment) and $\tilde{\beta}$ when $\delta = 1$ (observables as important as unobservables). The results in col. 11 and 12 are related to the full model presented in col. 10.

variability is small), Oster (2019) argues that the likelihood of an omitted variable problem due to self-selection on unobservables can be assessed by looking at the stability of the estimates across specifications: small variations across specifications is the sign of a low risk of estimation bias. Our R^2 is close to 1 when fixed effects are included (see col. 8 in Table

1.A23(b)). Col. 9 and 10 show that our coefficient of interest is remarkably stable when we include the other covariates. This suggests that the omitted variable bias is small. In col. 11, we push the methodology of Oster (2019) further and compute the relative degree of selection on unobservables to observables (δ). This ratio proxies the magnitude of the selection bias due to unobservables (compared to the bias due to observables) required to make the actual diversity coefficient statistically insignificant.[32] Selection on unobservables is more likely to be problematic if δ is small. In our context, δ is around 3, which means that selection on unobservables is unlikely to drive our results. Finally, in col. 12, we compute the bounding set of the coefficient of diversity under a fully identified model and a correction for selection on unobservables with $\delta = 1$. The bounding set of the coefficient of diversity never includes zero. This means that our coefficient of interest should remain positive, robust and significantly greater than the coefficient of the low skilled even if a reasonably low selection bias due to unobservables exists.[33]

[32]In Stata, we use the **psacalc** command developed by Oster (2019) to compute the relative degree of selection δ and the identification set. δ is the ratio between the covariance between diversity and unobserved controls to the covariance between diversity and the observed controls that generates a coefficient equal to 0.

[33]The same findings are obtained when adding the self-selection index computed in Eq. (1.12) and its interaction with diversity to the set of controls.

1.6.5 System GMM regressions

We provide results of dynamic GMM regressions with internal or external instruments in Table 1.A24 and Table 1.A25, respectively. These estimate allow us to jointly account for the endogeneity of the variable of interest, of the immigrant shares, and of other covariates. We use their own lags as instruments, with different lag structures. In these regressions, the lagged dependent is insignificant or weakly significant, which reinforces the credibility of our static benchmark specification. In addition, the effect of diversity is similar to that obtained in the static model.

Table 1.A24: System GMM: Internal instruments
Results by skill group (Dep= $log(y_{s,t})$)

	(1) $K=H$	(2) $K=H$	(3) $K=H$	(4) $K=H$	(5) $K=H$	(6) $K=L$	(7) $K=L$	(8) $K=L$	(9) $K=L$	(10) $K=L$
$log(y_{s,t-10})$	0.546***	0.269	0.307*	0.314*	0.309*	0.555***	0.293*	0.338*	0.339*	0.370*
	(0.116)	(0.190)	(0.179)	(0.168)	(0.174)	(0.189)	(0.172)	(0.188)	(0.185)	(0.188)
$MD^K_{s,t}$	1.792***	1.295**	1.173**	1.088**	1.231***	0.238*	0.371***	0.358***	0.362***	0.385***
	(0.527)	(0.580)	(0.467)	(0.448)	(0.374)	(0.119)	(0.103)	(0.102)	(0.103)	(0.104)
$m^K_{s,t}$	1.325***	0.892	0.896*	0.862	0.953*	0.313	0.147	0.357	0.541*	0.574*
	(0.483)	(0.563)	(0.485)	(0.519)	(0.509)	(0.324)	(0.363)	(0.401)	(0.318)	(0.321)
$log(Pop_{s,t})$	-0.065	-0.101	-0.082	-0.082	-0.083	-0.039	-0.093*	-0.067	-0.074	-0.062
	(0.046)	(0.072)	(0.061)	(0.051)	(0.059)	(0.044)	(0.053)	(0.054)	(0.049)	(0.050)
$log(Urb_{s,t})$	0.017	0.416	0.304	0.323	0.306	0.204	0.593**	0.397	0.358	0.280
	(0.125)	(0.431)	(0.316)	(0.309)	(0.255)	(0.145)	(0.277)	(0.266)	(0.258)	(0.203)
$log(Hum_{s,t})$	0.563**	0.356	0.439	0.395	0.465	0.359	-0.148	0.095	0.315	0.302
	(0.260)	(0.436)	(0.344)	(0.321)	(0.334)	(0.266)	(0.442)	(0.416)	(0.287)	(0.324)
Constant	2.979**	5.498***	5.261***	5.280***	5.133***	2.888*	4.895***	4.567***	4.445***	4.364**
	(1.184)	(1.897)	(1.794)	(1.514)	(1.748)	(1.583)	(1.173)	(1.365)	(1.184)	(1.650)
Observations	255	255	255	255	255	255	255	255	255	255
Nb. states	51	51	51	51	51	51	51	51	51	51
Nb. instruments	56	26	31	36	41	56	26	31	36	41
Nb. lags (endogenous var.)	2	2	3	4	5	2	2	3	4	5
Collapsed matrix	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AR(1)	0.006	0.021	0.020	0.015	0.019	0.006	0.012	0.011	0.012	0.010
AR(2)	0.134	0.118	0.121	0.126	0.124	0.256	0.200	0.210	0.215	0.222
Hansen J (p-value)	0.554	0.056	0.158	0.218	0.304	0.505	0.202	0.168	0.244	0.170

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. Source: Authors' elaboration on IPUMS-US data. The lagged dependent variable is always treated as predetermined and instrumented with its own first to second lags. $MD^K_{s,t}$; $m^K_{s,t}$; $log(Pop_{s,t})$; $log(Urb_{s,t})$ and $log(Hum_{s,t})$ are treated as endogenous variables and instrumented with their own first to X lags. The number of lags X is reported in the table. From columns (2) to (5) and (7) to (10) the matrix of endogenous variable is collapsed in order to keep the number of instruments below the number of states.

Table 1.A25: System GMM: External instruments
Results by skill group (Dep= $log(y_{s,t})$)

	(1) $K=H$	(2) $K=H$	(3) $K=H$	(4) $K=H$	(5) $K=L$	(6) $K=L$	(7) $K=L$	(8) $K=L$
$log(y_{s,t-10})$	0.184	0.215	0.202	0.228	0.293	0.361*	0.308*	0.357**
	(0.166)	(0.144)	(0.169)	(0.153)	(0.197)	(0.201)	(0.174)	(0.171)
$MD^K_{s,t}$	1.504**	1.587***	1.675***	1.791***	0.254*	0.263*	0.435***	0.446***
	(0.686)	(0.521)	(0.466)	(0.486)	(0.131)	(0.146)	(0.116)	(0.119)
$m^K_{s,t}$	1.353*	1.543**	1.435**	1.660**	0.266	0.546	0.416	0.655
	(0.684)	(0.618)	(0.603)	(0.620)	(0.388)	(0.416)	(0.433)	(0.450)
$log(Pop_{s,t})$	-0.132**	-0.119***	-0.128*	-0.122**	-0.088	-0.069	-0.092*	-0.080*
	(0.059)	(0.044)	(0.064)	(0.052)	(0.067)	(0.058)	(0.048)	(0.042)
$log(Urb_{s,t})$	0.382	0.267	0.336	0.242	0.474	0.280	0.472*	0.348
	(0.416)	(0.298)	(0.367)	(0.291)	(0.339)	(0.300)	(0.264)	(0.237)
$log(Hum_{s,t})$	0.349	0.503	0.332	0.502	0.029	0.280	-0.215	0.054
	(0.476)	(0.407)	(0.440)	(0.392)	(0.325)	(0.263)	(0.371)	(0.316)
Constant	6.729***	6.336***	6.545***	6.143***	5.115***	4.672***	5.320***	4.811***
	(1.775)	(1.435)	(1.667)	(1.323)	(1.347)	(1.456)	(1.333)	(1.306)
Observations	255	255	255	255	255	255	255	255
Nb. states	51	51	51	51	51	51	51	51
Nb. instruments	25	33	25	33	25	33	25	33
Nb. lags (endogenous var.)	2	4	2	4	2	4	2	4
Collapsed matrix	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AR(1)	0.032	0.023	0.036	0.029	0.016	0.014	0.014	0.008
AR(2)	0.097	0.104	0.097	0.102	0.190	0.206	0.188	0.213
Hansen J (p-value)	0.189	0.306	0.202	0.272	0.145	0.189	0.232	0.303

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the state level. Source: Authors' elaboration on IPUMS-US data. The lagged dependent variable is always treated as predetermined and instrumented with its own first to second lags. $m^K_{s,t}$; $log(Pop_{s,t})$; $log(Urb_{s,t})$ and $log(Hum_{s,t})$ are treated as endogenous variables and instrumented with their own first to X lags. The number of lags X is reported in the table. The matrix of endogenous variables is collapsed in order to keep the number of instruments below the number of states. In columns (1), (2), (5) and (6), $MD^K_{s,t}$ is instrumented using the augmented shift-share strategy while in columns (3), (4), (7) and (8), $MD^K_{s,t}$ is instrumented using predictions of the gravity-like strategy *a la* Feyrer (2019).

1.6.6 Adjusted diversity index

In the benchmark specification, the birthplace diversity index $MD^K_{s,t}$ does not account for the cultural distance between origin and destination countries. It assumes that all groups are culturally equidistant from each other. Here, we report the results obtained when the standard index of birthplace diversity is replaced by an adjusted index that accounts for the genetic distance between the countries of origin of immigrants. This adjusted index multiplies the probability that two randomly-drawn immigrants were born in two different countries by a measure of cultural distance between these two countries. For the latter, we use the database on genetic distance between countries, constructed by Spolaore and Wacziarg (2015). Genetic distance is based on blood sample and proxies the time since two populations had common ancestors. Spolaore and Wacziarg (2015) find a pattern of positive and significant relationships between genetic distance and various measures of cultural distance, including language, religion, values, and norms. We thus use the adjusted index of cultural diversity to investigate whether the variety effect is associated with the genetic distance between countries of origin.

The adjusted diversity index writes as:

$$MD^{K,G}_{s,t} = \sum_{i \neq s}^{I} \widehat{k}^K_{i,s,t} \sum_{j \neq i,s}^{J} \widehat{k}^K_{j,s,t} d^G_{i,j}, \tag{1.13}$$

where $d^G_{i,j} \in [0,1]$ is a normalized genetic distance between population from country i and country j. The correlation between the adjusted and the unweighted diversity indices is equal to 0.69 for the college-educated population, and to 0.38 for the less educated. The OLS-FE estimates point in the same direction as our benchmark regressions. The magnitude and the significance of the estimates are not statistically different from the ones reported in the previous tables. Using the gravity-like IV strategy, the effect of high-skilled diversity remains positive and significant. We lose significance under the adjusted shift-share IV strategy. However, the shift-share statistics are less convincing when we introduce more complexity in the diversity index. As far as low-skilled diversity is concerned, the effect remains insignificant in the OLS-FE and IV regressions. Overall, accounting for cultural distance between countries does not add much compared with the benchmark model.

Table 1.A26: Robustness to the measure of diversity
Accounting for genetic distance between countries (Dep= $log(y_{s,t})$)

	(1) FE OLS $K=H$	(2) 2SLS Shift-Share $K=H$	(3) 2SLS Feyrer $K=H$	(4) FE OLS $K=L$	(5) 2SLS Shift-Share $K=L$	(6) 2SLS Feyrer $K=L$
$MD_{s,t}^{S,G}$	0.658***	0.558	1.092***	0.199	0.363	0.438**
	(0.201)	(0.360)	(0.223)	(0.135)	(0.225)	(0.180)
$m_{s,t}^{K}$	0.298	0.287	0.346	0.413	0.479*	0.509*
	(0.326)	(0.299)	(0.346)	(0.284)	(0.260)	(0.291)
$log(Pop_{s,t})$	-0.134	-0.135*	-0.128	-0.143	-0.132	-0.127
	(0.080)	(0.079)	(0.080)	(0.088)	(0.092)	(0.092)
$log(Urb_{s,t})$	0.277*	0.278*	0.269*	0.284*	0.289*	0.291*
	(0.153)	(0.151)	(0.144)	(0.169)	(0.164)	(0.166)
$log(Hum_{s,t})$	0.818***	0.820***	0.808***	0.857***	0.816***	0.797***
	(0.207)	(0.198)	(0.214)	(0.206)	(0.208)	(0.201)
Constant	7.436***			7.547***		
	(1.310)			(1.349)		
Observations	306	306	306	306	306	306
Nb. states	51	51	51	51	51	51
R-squared	0.993	0.993	0.993	0.993	0.993	0.993
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
States fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
K-P F-$Test$	N.A	30.90	140.6	N.A	76.63	241.0
Stock Yogo		29.18/16.23	16.38/8.96		29.18/16.23	16.38/8.96
Hansen J (p-value)	N.A	0.388	N.A	N.A	0.232	N.A

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the State level. The specification is described in Eq.(3) and includes all fixed effects. OLS-FE results are provided in col. 1 and 4 and 5; IV results are provided in col. 2, 3, 5 and 6 using the augmented shift-share and gravity-like strategies. Results for college-educated migrants are provided in col. 1 to 3; results for the low-skilled are provided in col. 3 to 6. †Kleinbergenn-Paap F-statistic tests for weak identification (critical values from Stock-Yogo (2005) are given for 10%/15% maximal IV size). The sample includes the 50 US states and the District of Columbia from 1960 to 2010. The set of control variables includes the immigration rate ($m_{s,t}^{K}$), the log of population ($log(Pop_{s,t})$), the log of urbanization ($log(Urb_{s,t})$) and the log of the average educational attainment of the working-age population ($log(Hum_{s,t})$).

1.6.7 Analysis for the OECD countries

This appendix presents the results of a complementary analysis conducted on the 34 member states of the OECD. Despite the drawbacks of the OECD data (e.g. absence of information on the skill composition of migration stocks), we only want to verify whether the results obtained for the US states are not invalidated when using cross-country data, as the US is usually considered as one of the most attractive countries for (high-skilled) migrants.

Population data at the country level for the OECD member states are available from the Global Migrant Stock database described in Özden*et al* (2011). This database documents the bilateral stocks and shares of international migrants, $k^A_{i,s,t}$, in the population of each OECD country r, by country of origin i, and by year t. Özden*et al* (2011) collected and harmonized over 1,000 censuses and population registers to construct comprehensive matrices of origin-destination stocks that correspond to the last five completed census rounds, i.e. for the period 1960-2000 in 10-year intervals. They specified a standard and common set of countries for the entire period, disaggregating data for the countries that no longer exist on the basis of more recent migration figures. There is no artificial variation due to the dislocation of the Eastern Block. We expanded the database by adding the share of native citizens, $k^A_{r,r,t}$, in order to match the total population data. We also added the year 2010 using the bilateral stock estimates of the United nations (2013).[34] Hence, our OECD database covers the same period (1960-2010) in ten-year intervals as the IPUMS data; results for the US states and for OECD countries are then comparable. Still, compared to the IPUMS data, the OECD data suffer from two drawbacks. The first drawback is that it does not report the educational structure of migration stocks. The second drawback is that many imputations were used to fill the missing bilateral cells.

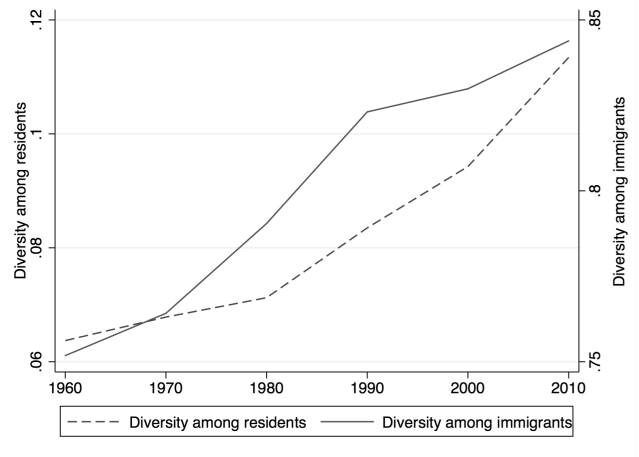
We compute our indices of birthplace diversity, $TD^A_{s,t}$ and $MD^A_{s,t}$, for each OECD member state. Figure 1.A6 shows that most OECD member states have experienced increasing immigration rates and total diversity indices ($TD^A_{s,t}$) in the aftermath of WW2. This is particularly the case after the year 1980.[35] Compared to the US, the average level of diversity among immigrants ($MD^A_{s,t}$) increased more strongly. Unsurprisingly, this average trend conceals important disparities across countries. Although a rise in the variety of immigrants was observed in most countries, diversity decreased in countries such as the US, Mexico or Slovakia.

Results of the OECD regressions are depicted in Table 1.A28. The first three columns of the table show the effect of birthplace diversity among immigrants on GDP per capita, regardless

[34]The list of the 34 OECD member states as well as descriptive statistics are available in Table 1.A27.
[35]Diversity trends for OECD countries are described in Figure 1.A7.

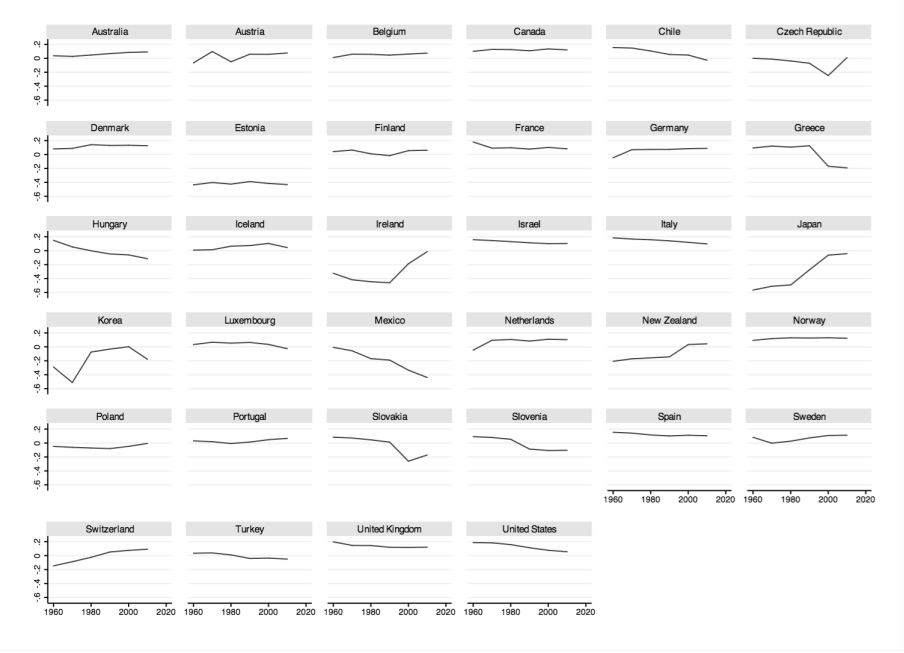
of the educational structure. It is worth noticing that these estimates include the log ratio of trade to GDP, the logarithm of the population, the log of the number of years of schooling in the working-age population and the *Polity2* index of democracy as covariates. As in the US state sample, the effect of birthplace diversity on GDP per capita is strongly positive and significant at the 1% level when using pooled OLS, OLS-FE and the gravity-like IV strategy *a la* Feyrer (2019)). The magnitude of the effect is larger than in the US sample: a one standard deviation change in birthplace diversity is associated with a 13% increase in GDP per capita. This implies that the Japanese level of GDP per capita would be 1,770 dollars greater if Japan had the same diversity index as the US in 2010.

Figure 1.A6: Trends in birthplace diversity in the OECD member states, 1960-2010



Notes: Diversity among residents is defined as in Eq. (1.1), whereas diversity among immigrants is defined as in Eq. (1.2). Source: Authors' elaboration on Özden *et al* (2011).

Figure 1.A7: Diversity among immigrants ($MD^A_{s,t}$) in the OECD countries



Source: Authors' elaboration on Özden*et al* (2011) data.

Source: Authors' elaboration on Özden*et al* (2011). Diversity among residents is defined as in Eq. (1). Diversity among immigrants is defined as in Eq. (2).

In the remainder of the table, we combine the data of Özden*et al* (2011) with cross-sectional data on the skill structure of migration stocks. We use the database of Artuc *et al* (2015), which documents the proportion on college-educated immigrants in all OECD countries for the years 1990 and 2000. In col. 4, we add an interaction term, the product of the average birthplace diversity index by the proportion of college-educated immigrants observed in 2000. Despite collinearity with the non-interacted index, the interaction term is positive and significant, whereas the average index of diversity looses significance. This suggests that the effect of birthplace diversity increases with the educational level of migrants. At the median level of the proportion of college graduates (22%), the marginal effect of birthplace diversity on GDP per capita is significant at the 1% level, and the coefficient is almost equal to that obtained with the US state sample (0.506). The results of col. 4 are illustrated on Figure 1.A8. Interestingly, this figure shows that the effect of birthplace diversity is insignificant when the bilateral migration stocks is mainly composed of low-skilled migrants. However, above 20% of college graduates, the effect of diversity becomes significant, and increases with the proportion of college graduates. Again, this suggests that the effect of birthplace

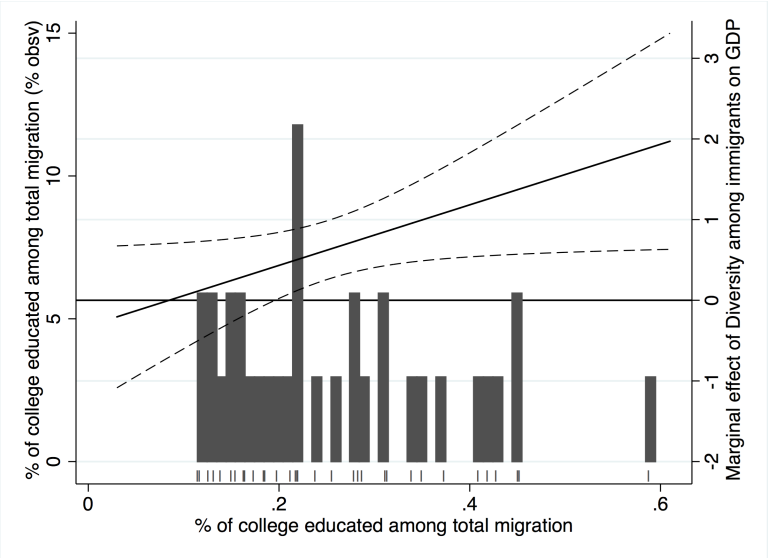
Table 1.A27: Descriptives statistics for the OECD countries (34)

	$log(y_{s,t})$	$MD^A_{s,t}$	$m^A_{s,t}$	$log(Pop_{s,t})$	$log(Trade_{s,t})$	$Democ_{s,t}$	$log(Hum_{s,t})$
Australia	9.655	0.860	0.202	16.563	3.486	1.000	2.338
Austria	9.543	0.830	0.113	15.855	4.226	1.000	2.095
Belgium	9.576	0.852	0.081	16.114	4.699	0.983	2.151
Canada	9.695	0.920	0.168	17.053	3.948	1.000	2.298
Chile	8.843	0.881	0.012	16.294	3.851	0.758	1.958
Czech Republic	9.018	0.741	0.035	16.132	4.349	0.542	2.366
Denmark	9.690	0.917	0.044	15.445	4.266	1.000	2.046
Estonia	9.254	0.384	0.194	14.134	4.739	0.475	2.201
Finland	9.516	0.836	0.018	15.401	4.047	1.000	1.941
France	9.594	0.905	0.099	17.816	3.708	0.900	1.901
Germany	9.544	0.857	0.109	18.194	3.814	0.763	2.193
Greece	9.034	0.814	0.038	16.096	3.643	0.792	2.024
Hungary	8.713	0.797	0.038	16.146	4.575	0.575	2.213
Iceland	9.561	0.852	0.045	12.384	4.361	1.000	2.064
Ireland	9.290	0.492	0.071	15.057	4.685	1.000	2.186
Israel	9.286	0.926	0.395	15.217	4.169	0.975	2.305
Italy	9.465	0.945	0.035	17.831	3.646	1.000	1.898
Japan	9.465	0.476	0.010	18.550	3.121	1.000	2.191
Korea	8.607	0.620	0.003	17.443	3.881	0.708	1.965
Luxembourg	9.896	0.838	0.248	12.863	5.309	1.000	2.029
Mexico	8.628	0.601	0.005	18.107	3.459	0.508	1.512
Netherlands	9.647	0.875	0.061	16.468	4.682	1.000	2.220
New Zealand	9.492	0.700	0.162	14.998	4.047	1.000	2.396
Norway	9.655	0.921	0.049	15.244	4.299	1.000	2.256
Poland	8.632	0.748	0.043	17.379	3.973	0.533	2.114
Portugal	8.999	0.829	0.038	16.083	3.978	0.675	1.405
Slovakia	8.911	0.764	0.011	15.412	4.335	0.550	2.340
Slovenia	9.221	0.789	0.054	14.429	4.606	0.467	2.228
Spain	9.142	0.923	0.040	17.440	3.510	0.708	1.820
Sweden	9.666	0.868	0.088	15.946	4.101	1.000	2.263
Switzerland	9.855	0.794	0.197	15.691	4.398	1.000	2.352
Turkey	8.402	0.793	0.024	17.659	3.005	0.775	1.210
United Kingdom	9.587	0.942	0.066	17.859	3.885	1.000	2.162
United States	9.906	0.929	0.085	19.307	2.855	1.000	2.449

Source: Authors' elaboration on Özden *et al* (2011).

diversity on macroeconomic performance is skill-specific. In the two last columns of Table 1.A28, we tentatively proxy the stocks of high-skilled and low-skilled migrants using the recent shares of college-educated migrants provided in Artuc *et al* (2015). We use the 1990 skill shares to split the bilateral migration stocks observed in 1960, 1970, 1980 and 1990; we use the 2000 skill shares to split the stocks observed in 2000 and 2010. We find a positive and significant effect of birthplace diversity in both regressions, and a greater effect for high-skilled diversity. Although we should not give too much credit to these results, we confirm that the patterns obtained for the US states are not invalidated when using cross-country data.

Figure 1.A8: Marginal effect of diversity conditional to the share of college graduates in the 2000 immigrant stock



Notes: The solid line is based on col. 4 of Table 1.A28. It represents the marginal effect of birthplace diversity on GDP per capita conditional on the percentage of college graduates in the total stock of bilateral migrants in 2000. The histogram indicates the percentage of observations of the modifying variable, and each mark on the vertical axis represents one OECD country. The dashed line depicts the upper and lower bounds of the 95% confidence interval.

Table 1.A28: Birthplace diversity in a cross-country setting.
Regressions for the OECD member states (Dep= $log(y_{s,t})$)

	(1) OLS $S=A$	(2) FE $S=A$	(3) 2SLS $S=A$	(4) 2SLS $S=A$	(5) 2SLS $K=H$	(6) 2SLS $K=L$
$MD^K_{s,t}$	0.834***	0.788**	0.940***	-0.318	1.016***	0.901***
	(0.274)	(0.292)	(0.245)	(0.500)	(0.316)	(0.259)
$MD^A_{s,t} \times CollMig_{s,t}$				3.754**		
				(1.839)		
$m^K_{s,t}$	0.683	-0.291	0.148	0.020	0.107	-0.096
	(0.911)	(0.333)	(0.429)	(0.374)	(0.070)	(0.172)
$log(Pop_{s,t})$	-0.002	0.005	0.076	-0.001	0.073	0.025
	(0.027)	(0.144)	(0.143)	(0.138)	(0.145)	(0.152)
$log(Trade_{s,t})$	0.073	0.150	0.157*	0.146*	0.140	0.160*
	(0.099)	(0.089)	(0.086)	(0.080)	(0.092)	(0.083)
$log(Hum_{s,t})$	0.507***	0.370*	0.363**	0.425**	0.425**	0.347**
	(0.152)	(0.187)	(0.181)	(0.181)	(0.196)	(0.170)
$Democ_{s,t}$	0.578***	0.039	0.076	0.046	0.039	0.068
	(0.105)	(0.078)	(0.070)	(0.071)	(0.068)	(0.077)
Constant	6.420***	6.750**				
	(0.702)	(2.625)				
Total effect of $MD^A_{s,t}$				0.506***		
				(0.192)		
Observations	204	204	204	204	204	204
Nb. countries	34	34	34	34	34	34
R-squared	0.750	0.901	0.899	0.908	0.897	0.903
Time fixed effects	No	Yes	Yes	Yes	Yes	Yes
Country fixed effects	No	Yes	Yes	Yes	Yes	Yes
K-P F-$Test^{\dagger}$			40.89	92.66	31.64	35.18
Stock Yogo			7.03/4.58	13.43/8.18	7.03/4.58	7.03/4.58

Notes. *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses are clustered at the country level. The $CollMig_{s,t}$ variable is the share of college-educated migrants in the total stock of immigrants of the receiving country in 2000. The total effect of diversity in col. (4) is computed for $CollMig_{s,t}$ at its median level. Diversity and immigration rates are instrumented using the gravity-like IV strategy *a la* Feyrer (2019). Zero-stage estimates are available in the appendix (Table 1.A20). †Kleinbergenn-Paap F-statistic tests for weak identification (critical values from Stock-Yogo (2005) are given for 10%/15% maximal IV size). The sample includes the 34 OECD member states from 1960 to 2010. The set of control variables includes the immigration rate ($m^K_{s,t}$), the log of population ($log(Pop_{s,t})$), the log of trade ($log(Trade_{s,t})$), the log of the number of years of schooling in the working-age population ($log(Hum_{s,t})$), and the Polity2 index of democracy ($Democ_{s,t}$).

Table 1.A29: Zero-stage estimates (PPML): gravity model *a la* Feyrer (2019)

	(1)
	All
	OECD
	$log(Stock_{i,s,t})$
$log(Dist_{i,s}) \times I_{1960}$	-0.816***
	(0.155)
$log(Dist_{i,s}) \times I_{1970}$	-0.876***
	(0.151)
$log(Dist_{i,s}) \times I_{1980}$	-0.774***
	(0.139)
$log(Dist_{i,s}) \times I_{1990}$	-0.683***
	(0.131)
$log(Dist_{i,s}) \times I_{2000}$	-0.617***
	(0.125)
$log(Dist_{i,s}) \times I_{2010}$	-0.573***
	(0.124)
$Bord_{i,s}$	0.755***
	(0.219)
$Lang_{i,s}$	1.565***
	(0.218)
Constant	15.078***
	(1.322)
Observations	38964
Nb. origin	191
Nb. destination	34
R-squared	0.727
Year dummies	Yes
Origin dummies	Yes
Destination dummies	Yes

Notes. *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses adjusted for clustering at the state/country-pair level. Distances data are not available for Liechtenstein, Luxembourg and Holy See.

Chapter 2

Does Birthplace Diversity affect Economic Complexity? Cross-country Evidence

Abstract[1]

We empirically investigate the relationship between a country's economic complexity and the diversity in the birthplaces of its immigrants. Our cross-country analysis suggests that birthplace diversity is strongly and positively associated with economic complexity. This holds particularly for diversity among highly educated migrants and for countries at intermediate levels of economic complexity. The results are robust to accounting for previous trends in birthplace diversity as well as to using alternatives diversity measures. We address endogeneity concerns by instrumenting diversity through predicted stocks from a pseudo-gravity model as well as from a standard shift-share approach. Finally, we provide evidence suggesting that birthplace diversity boosts economic complexity by increasing the diversification of the host country's export basket.

2.1 Introduction

Diversity of immigration (or immigrants' birthplace diversity) has been shown to boost economic performance at different levels of aggregation: countries (Ortega and Peri (2014),

[1]This paper is coauthored with Dany Bahar and Hillel Rapoport.

Alesina et al. (2016) and Bove and Elia (2016)), regions (Trax *et al.* (2015)), US metropolitan areas, counties and states (Ottaviano and Peri (2006), Ager and Brückner (2013), Fulford et al. (2017), Docquier *et al.* (2019b)), or firms (e.g., Parrotta et al. (2014), Trax *et al.* (2015)).[2] The dominant interpretation for the positive association between birthplace diversity and economic performance has to do with the complementarity in skills and knowledge sets brought about by immigrants who grew up in different environments, went to different school systems, and learned different trades and skills as a result. As immigrants expand the set of skills a country can access to, by the same token they also expand its opportunities to become competitive in a broader set of economic activities. One way in which economists have recently measured the extent to which an economy is capable of becoming competitive in a more diverse set of industries is by quantifying its "economic complexity" (Hausmann and Klinger (2007), Hidalgo *et al.* (2007), Hidalgo and Hausmann (2009)). This paper empirically explores the relationship between a country's population birthplace diversity and its levels of economic complexity.

During the last decades, virtually all modern economies have become more demographically diverse. While the global share of immigrants has remained stable over time –at about three percent of the world population– the share of the foreign-born population in OECD countries has increased from seven to thirteen percent, on average, between 1970 and 2010. Among them, some countries experienced a dramatic growth of the share of foreign born over the 1970-2010 period. For example, Spain (from 1.1 to 15 percent), Greece (from 1 to 10 percent), and Portugal (from 1 to 9 percent).[3] Moreover, the composition of the foreign-born population in OECD countries has changed mostly due to immigration from developing countries, with the average share of non-OECD migrants growing by a factor of 8.7 over the 1970-2010 period against just 1.2 for OECD migrants over the same period.

Our focus on economic complexity is based on the findings of a burgeoning literature which has documented how more complex economies tend to enjoy higher levels of income and better growth prospects (e.g. Hidalgo and Hausmann (2009)). Exploring the different ways through which countries achieve higher complexity, however, is still an open question. A common interpretation of the drivers of economic complexity is the availability of accumulated unobserved capabilities, or productive know-how, which in turn is reflected in the compo-

[2] At very low levels of aggregation such as teams, there are many instances of negative outcomes, especially when focusing on ethnic diversity in contexts of conflict (e.g., Hjort (2014), Lyons (2017)). In this paper we focus on birthplace diversity, rather than other measures of diversity used in the literature, such as ethnic (Alesina and La Ferrara (2005)) or linguistic (Desmet *et al.* (2012)) diversity.

[3] The OECD countries with the highest share of foreign-born in 2010 are Israel (36 percent), Luxembourg (33 percent), Australia (25 percent) and Canada (21 percent), according to the World Bank's Bilateral Migration Database (available here).

sition of a country's export basket, namely its level of diversification and its uniqueness. Achieving an export basket that is more diversified and includes less ubiquitous products (and, hence, generates higher prospects of economic growth), conceptually, would imply drawing from a more diverse set of complementary skills in the economy. Following this line of thought, we investigate whether economic complexity is positively affected by higher birthplace diversity of the local labor force. In essence, the more diverse a country is in terms of its population's birthplaces, the more accumulated knowledge there is in terms of idiosyncratic individual characteristics to draw from. As long as these characteristics and skills are complementary one to the other (and to the native set of skills and knowledge), they should translate into higher levels of economic complexity, income per capita, and economic growth.

In this paper, we put these ideas to the test. In particular, we gather data on bilateral migration stocks between countries to construct measures of immigration diversity and investigate how these relate to measures of economic complexity (i.e., to the "Economic Complexity Index" of Hidalgo and Hausmann (2009), Hausmann and Hidalgo (2011)). Our baseline regressions use the Herfindahl index of diversity (a measure of the likelihood that two randomly drawn immigrants are from different countries), however in our robustness analysis we use alternative measures of diversity such as the Theil index and its components. We go beyond Alesina et al. (2016) as we look at a potential driver behind their results, showing that birthplace diversity affects complexity which in turns affects growth. Moreover, since economic complexity (as measured in Hidalgo and Hausmann (2009)) is a combination of countries' export diversification and abilities to produce more unique products, we are also able to decompose the relation between diversity and complexity, shedding light on the mechanisms through which diversity ultimately affects economic growth.

We find that, on average, countries with one standard deviation from our sample mean in terms of birthplace diversity exhibit higher levels of economic complexity by 0.18 standard deviations above the mean. These results are particularly strong when limiting our birthplace diversity measure to skilled migrants (i.e., to migrants with college education or more), consistently with the idea that the relationship between diversity and complexity is driven by skill-complementarity. Our results are robust to controlling for the diversification level of countries' export baskets and their income levels. In other words, countries with higher birthplace diversity of their populations tend to be more economically complex, regardless of their income level and the baseline composition of their export baskets. Furthermore our results are particularly strong among developing countries that are in the middle of the economic complexity distribution, suggesting that an increase in skill variety is particularly relevant for boosting countries development path. We also find that our results are particu-

larly strong for inflows of immigrants from origin countries with little to no prior immigration in the host country of interest.

We further show that our results are robust to using the Theil index of diversity and its decomposition into a between and a within component. We find that most of the effect is driven by the between component, which can be interpreted as the extensive margin of diversity in the sense that it captures the contribution of new origin countries to the evolution of immigration diversity. This is consistent with our interpretation that the relation between diversity and complexity is fundamentally about expanding the set of skills and knowledge to be combined in production. Our results are robust to controlling for potential omitted factors related to immigration diversity such as transfers of technological norms from origin to destination countries with an origin-specific effect (Valette (2018)), or to controlling for birthplace diversity of immigrants in 1960 (thereby ruling out the possibility that the results are driven by the diversity of the second and third generations of immigrants). Finally, we are able to rule out our results being driven by time-invariant unobserved heterogeneity when looking at a short panel of high income countries.[4] We also test the robustness of our cross-country analysis once we control for time-invariant country-specific factors, showing that our estimates are still robust and unlikely explained by missing unobserved factors.

In an attempt to go beyond correlation and deal with endogeneity concerns, we propose two instrumental variable approaches. The first is based on the prediction of skill-specific bilateral stocks of immigrants through a gravity model (see Alesina et al. (2016) and Docquier *et al.* (2019b)). Including a set of common gravity controls from the trade literature (such as distance, common language, sharing borders, etc.) and interactions between origin and year dummies to capture all the push-time variant factors, we then predict skill-specific bilateral stocks. With those predicted stocks we built our first set of instrumental variables, as skill-specific predicted birthplace diversity indexes. The second approach is based on the shift-share instrument (Card (2001) and Moriconi *et al.* (2018)) to predict the skill-specific bilateral immigration stocks based on the location of immigrants by origin countries in the 1970s and the aggregate skill-specific flows from origin countries. The predicted stocks are then determined by the previous "enclaves" of immigrants and not by recent economic/social factors. Therefore, we argue, our instruments based on this methodology are less prone to reverse causation bias. Moreover, since important historical and migration specific events happened which are orthogonal to the cross-country distribution of immigrants in the 1970s

[4]The only data source that provides skill-specific bilateral stocks of migration is the Artuc et al. (2015) dataset over the 1990-2000 period and the Database on Immigrants in OECD countries (DIOC) over the 2000-2010 period. Only high-income countries are available in both databases.

and produced exogenous variation of migration flows,[5] the 1970 distribution of immigrants can reasonably be less correlated with country persisting factors. Those exogenous shocks were indeed large and strengthen the reliability of our approach, since the correlation between our instrument and unobserved persistent countries factors is mitigated. Combining both approaches (gravity model and shift-share based instruments), our estimates support a causal interpretation for the effect of diversity on economic complexity, in particular among developing countries.

The rest of the paper is organized as follows. Section 2.2 discusses the data sources and presents some stylized facts on the evolution and distribution of Economic Complexity and birthplace diversity. Section 2.3 presents our empirical approach and our strategies to deal with endogeneity. Section 2.4 shows the main results of the analysis. In Section 2.5 we discuss the robustness of our results. Section 2.6 discusses candidate mechanisms. Finally, Section 2.7 concludes.

2.2 Data and Stylized Facts

Our paper investigates the relation between the birthplace diversity of immigrants and the level of economic complexity of the receiving countries. Combining different sources, our main analysis covers a sample of 100 countries over the period 1990 to 2000. In Section 2.2.1 we describe our data on skill-specific immigration and various measures of birthplace diversity, and in Section 2.2.2 we describe the measures of economic complexity used. In Section 2.2.3 we present some stylized facts that include suggestive evidence of a relationship between birthplace diversity and complexity, as a preamble to our econometric estimations.

2.2.1 Immigration and Diversity Indexes

Our data source on migration stocks comes from Artuc et al. (2015) (henceforth Artuc et al. (2015)). This dataset provides a square matrix of bilateral migration stocks for 195 countries for the years 1990 and 2000. It also contains migration stocks disaggregated by education level of migrants, splitting them between college and non-college educated, as well as disaggregated by gender. In the dataset, an immigrant is defined as a 25 year or

[5]Some examples are the fall of the Soviet Union and resulting opening of the Iron Curtain in 1989, the creation of the European Union and of the Schengen Area that allowed free movement of Europeans in 1995 and the recent implementation of the Hart-Cellar Act which changed the quota system in the US and became effective at the end of 1968.

older foreign-born individual. Due to the high coverage of destination and origin countries and the skill composition of the stocks, this data set is increasingly used in cross-country analysis (e.g., Docquier *et al.* (2015), Alesina et al. (2016), Bahar and Rapoport (2018)). In keeping with the literature on birthplace diversity reviewed in the introduction and following the decomposition of birthplace diversity over the total population suggested by Alesina et al. (2016), we compute a country-level fractionalization index based on the immigrant population by skill group.[6] Namely, given destination country d, origin country $o \in O = \{1, ..., 195\}$ and year t, our immigrant birthplace diversity index is:

$$Div_{d,t,s}^{Mig} = \sum_{o}^{O} \overline{m}_{d,o,t,s}(1 - \overline{m}_{d,o,t,s}) \tag{2.1}$$

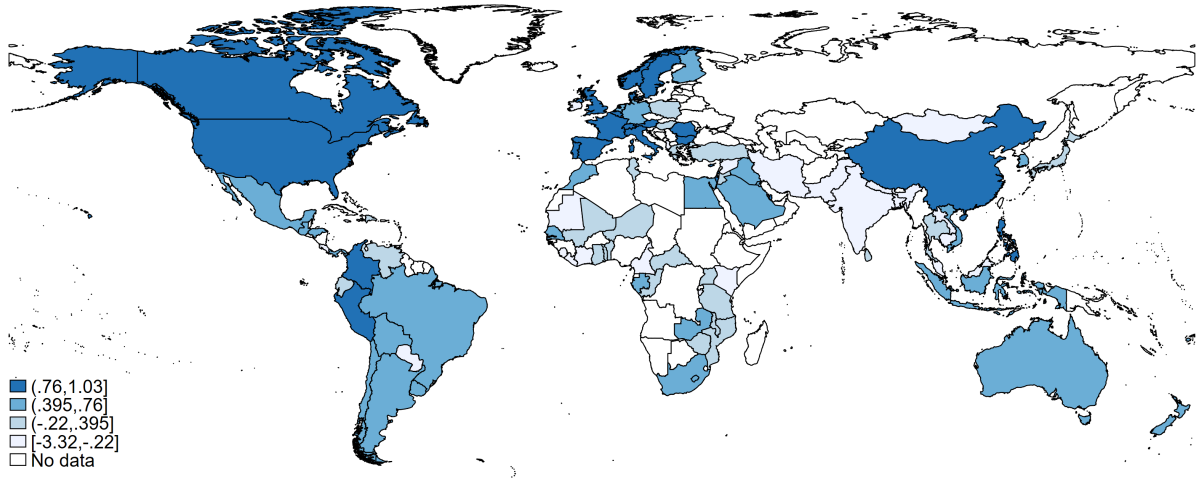
where $\overline{m}_{d,o,t,s}$ is the stock of immigrants from country of origin o in destination country d at year t with education $s \in E = \{All, HS, LS\}$ over the total stock of immigrants in country d with the same education level. Thank to the decomposition of the fractionalization index and including simultaneously in our empirical strategy both the immigration birthplace diversity index and the immigration share allows us to control both for the composition and the size of immigrants population. Moreover, as it is pointed out by the literature, the correlation between the immigration share and the same birthplace diversity index computed on the total country population rather then immigrant population is extremely high and close to one even in our sample (0.981).

By construction, our immigrant birthplace diversity index is between 0 and 1, and it measures the probability to draw randomly two individuals from the immigrants population that are born in different countries. A higher value for the index implies a more diverse immigrant population. To simplify further analysis, we standardized our measure of diversity with mean 0 and standard deviation equal to 1. Figure 2.1 plots the standardized level of birthplace diversity across countries. As can be seen, the birthplace diversity of immigrants can be quite high in countries with large foreign-born populations (e.g., in Western countries) or in countries where the share of foreign-born is small (e.g., China). In fact, there is little correlation between the share of immigrants in the host-country population and their diversity (see Alesina et al. (2016)).

Computing a skill-specific fractionalization index over the immigrants population is not

[6] Alesina et al. (2016) show that a birthplace fractionalization index computed over the total population can be decomposed in two components: the between component, which captures the diversity driven by the overall immigrants population and natives, and the within component, which captures the diversity within immigrants population. The correlation between the former component and the immigration share is extremely high and around 0.98.

Figure 2.1: Birthplace Migration Diversity 2000



Note: authors' calculations on Artuc et al. (2015) data. The map plots the standardized value of the birthplace diversity index on the overall population of migrants in year 2000.

the only way to measure countries' diversity. An alternative and opposite approach is to measure countries polarization rather than diversity. The intuition is that a more polarized population can be associated with a reduction of social cohesion and public good provision. Since the highest level of polarization index is associated with a population characterized by two homogeneous groups-of-origin immigrants, a highly polarized society implies less diversity,and so a less diverse set of skills and competences. Following Montalvo and Reynal-Querol (2005) and Ager and Brückner (2013) we compute a skill-specific polarization index over the immigrants population which captures how far the population distribution is from a bimodal distribution, which is written:

$$Pol^{Mig}_{d,t,s} = 1 - \sum_{o}^{O} \left(\frac{0.5 - \overline{m}_{d,o,t,s}}{0.5} \right)^2 \overline{m}_{d,o,t,s}. \tag{2.2}$$

Since we compute it over the immigrant population, this index reaches its maximum level when a country's immigrant population is composed by two groups of equal size. Table A1 shows that skill-specific fractionalization and polarization are negatively correlated. Moreover, if our prior is that the positive relation between population diversity and economic complexity is driven by the expansion of the set of skills and competences brought by different immigrants, we should find a negative relation between the birthplace polarization index and economic complexity. We will test this hypothesis and we will show it later on in our results.

Another alternative index of population diversity is the Theil index. Adopted in the trade literature (e.g., Cadot *et al.*, 2013), such index is used to compute measures export diversification using as inputs a vector of a country's per-product export shares. We can apply this index, too, to measure diversity by computing a skill-specific Theil index over the immigrant population as follow:

$$Theil^{Mig}_{d,t,s} = \frac{1}{N}\sum_{o}^{O}\left(\frac{MIG_{d,o,t,s}}{\mu_{d,t,s}}\right) ln\left(\frac{MIG_{d,o,t,s}}{\mu_{d,t,s}}\right) \tag{2.3}$$

where $MIG_{d,o,t,s}$ is the total stock of immigrants from origin country o in destination country d at year t with education s. The number of immigrants countries of origin is represented by $N = 195$, while $\mu_{d,t,s}$ is the average skill-specific size of immigrants group.[7] Compared to our benchmark fractionalization index computed in equation (2.1), the Theil index has two substantial differences. First, the Theil index is equal to zero when we have perfect diversification within the immigrants population, implying that immigrants are perfectly distributed among countries of origin. For this reason the Theil index is negatively related with our fractionalization index, as Table A1 shows. Second, the Theil index can be decomposed along two dimensions, allowing us a better understanding of the relation between diversity and economic complexity. Following Cadot *et al.* (2011), we decompose the Theil index in two additive components: the between-origin component ($Theil^{B,Mig}_{d,t,s}$) and within-origin component ($Theil^{W,Mig}_{d,t,s}$). The former captures the extensive margin of the Theil Index, which implies a variation in the number of origin countries represented in the immigrants population, while the latter capture the intensive margin, which is driven by the change of balance among immigrants groups. Analyzing the relation between the Theil index (and its components) and economic complexity can help decompose the relationship we are studying by looking at whether economic complexity correlates differently with a more diverse migrant population in terms of its composition (e.g., the distribution of the shares of each group) or, rather, the number of origin countries the migrants are coming from.

2.2.2 The Economic Complexity Index

Measuring a country's capabilities to produce and export, and quantify its impact on future economic growth, is challenging. Measures of productivity, based on the seminal Solow residual (Solow, 1956) to more advanced methodologies that incorporate socioeconomic indicators to measure countries' growth capacity are still the subject of numerous studies.

[7] Namely, we compute $\mu_{d,t,s}$ as follow: $\mu_{d,t,s} = \frac{1}{N}\sum_{o}^{O} MIG_{d,o,t,s}$.

Our paper focuses on one of those measures that has been at the core of a burgeoning literature related to economic growth: the Economic Complexity Index (ECI), developed by Hidalgo and Hausmann (2009). The ECI aims to capture a country's accumulated capabilities or productive know-how. Countries with highly diversified export baskets which in turn include industries exported by fewer countries have higher levels of economic complexity.Hidalgo and Hausmann (2009) document that higher levels of economic complexity positively relate to income per capita it has a particularly strong explanatory power in predicting future economic growth (as compared to other long-established determinants of economic growth).

For our exercise we collect yearly ECI data for 222 countries from 1962 to 2016, from the Atlas of Economic Complexity (Hausmann *et al.* (2014)).[8] As explained by Hausmann *et al.* (2014), the ECI is constructed using exports data for each country and year on nearly 800 tradable industries categorized using the SITC 4-digits classification.

As explained in Hidalgo and Hausmann (2009), the ECI is computed using product level export value data as its main input. It organizes the export data for every year in the form of a matrix M_{dp}, sized $d \times p$, where a cell is 1 if country d exports product p with comparative advantage, and 0 otherwise. To quantify whether a country exports a product with comparative advantage, the authors rely on the measure Revealed Comparative Advantage (RCA), originally suggested by Balassa (1965). The RCA for country d and product p for any year is computed according to the following formula:

$$RCA_{d,p} \equiv \frac{exp_{d,p} \sum_{p} exp_{d,p}}{\sum_{d} exp_{d,p} \sum_{d} \sum_{p} exp_{d,p}}$$

where $exp_{d,p}$ is the exported value of product p by country d. Thus, to create the M_{dp} matrix, it is assumed that country d exports product p with comparative advantage if its RCA is 1 or more. This threshold can be interpreted as that country exporting that product in higher relative proportion than the World as a whole.

As a first step to compute ECI, M_{dp} is used to quantify two indicators. First, for each country, the diversification of a country's export basket, measured as the number of products in which the country holds a RCA equal or greater than 1. Second, for each product, the number of countries that export such product with a RCA of unit or more. The indexes of country diversification ($K_{d,0}$) and product ubiquity ($K_{p,0}$) are defined as follow:

[8] Data are available from `atlas.cid.harvard.edu`

$$K_{d,0} = \sum_{p} M_{dp} \tag{2.4}$$

$$K_{p,0} = \sum_{d} M_{dp} \tag{2.5}$$

Equations (2.4) and (2.5) are the building blocks of the measure of economic complexity: countries with a more diverse export basket of less ubiquitous products are characterized by a high degree of economic complexity. The intuition is that both aspects (diversification and less ubiquitous products) requires more know-how and capabilities (thus, countries export more varieties) that, in turn, are more exclusive and rare (thus, the varieties a country exports, on average, are less common).

The calculation of the ECI as explained by Hidalgo and Hausmann (2009) relies on the application of what they define as the *Method of Reflections*. The construction of ECI is an iterative process. The method starts with $K_{d,0}$ in its first iteration, as a measure of a country's complexity simply by counting the number of products it exports (in a given year). Then, it incorporates information of each one of those products using $K_{p,0}$, and looks at the number of products d exports weighted by the ubiquity of each one of them. It then incorporates the average diversification of the countries that export the same products as d, and iterates again and again, until converging. As explained by Hidalgo and Hausmann (2009), the nth iteration of this measure is:

$$K_{d,n} = \frac{1}{K_{d,0}} \sum_{p} M_{dp} \times K_{p,n-1} \tag{2.6}$$

Where,

$$K_{p,n-1} = \frac{1}{K_{p,0}} \sum_{d} M_{dp} \times K_{d,n-2} \tag{2.7}$$

Substituting (2.7) in (2.6), we have:

$$K_{d,n} = \frac{1}{K_{d,0}} \sum_{p} M_{dp} \times \frac{1}{K_{p,0}} \sum_{d'} M_{d'p} \times K_{d',n-2} \tag{2.8}$$

As shown by Hausmann *et al.* (2014), this expression can be rewritten as:

$$K_{d,n} = \sum_{d'} \widetilde{M_{dd'}} \times K_{d',n-2} \tag{2.9}$$

Where,

$$\widetilde{M_{dd'}} = \sum_{p} \frac{M_{d,p} M_{d',p}}{K_{d,0} K_{p,0}} \tag{2.10}$$

Hausmann *et al.* (2014) note that equation (2.9) is satisfied when $K_{d,n} = K_{d,n-2}$ (e.g., convergence is achieved). Using matricial algebra, this corresponds to the eigenvector of $M_{dd'}$ associated with the second largest eigenvalue (where captures most of the variance). Following their notation, let's call this eigenvector $\vec{K}$. Thus, the vector of all ECI values (one per country) is computed by standardizing $\vec{K}$:

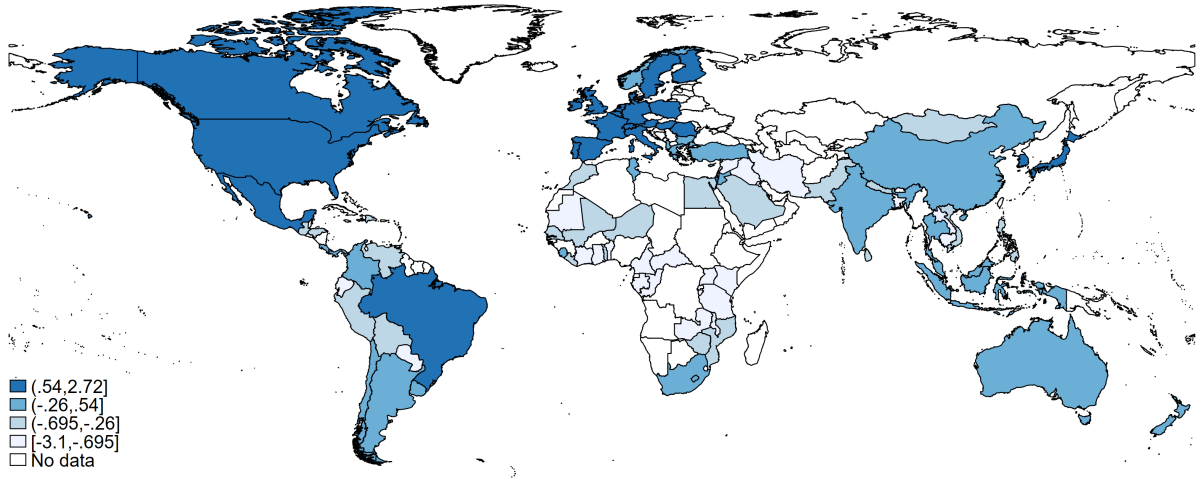
$$ECI = \frac{\vec{K} - < \vec{K} >}{stdev(\vec{K})} \tag{2.11}$$

Where $< . >$ and $stdev(.)$ represent average and standard deviation, respectively.

This calculation provides one ECI value for each country d in every year, using data on its export basket.

Figure 2.2 plots the distribution of ECI values per country for year 2000. We can clearly see that the geographic distribution of the intensity in economic complexity is rather similar to the one of birthplace diversity. Moreover, developed high-income societies are characterized by an high level of economic complexity, compared to developing/low-income countries.

Figure 2.2: Economic Complexity Index 2000



Note: authors' calculations using data from Hausmann *et al.* (2014). The map plot the standardized value of the Economic Complexity Index by country in year 2000.

2.2.3 Stylized facts

While there is a literature that has focused its attention to investigating the relation between birthplace diversity and several economic outcomes, such as income per capita and productivity, our focus is particularly on economic complexity for three main reasons.

First, we suspect that the measure of economic complexity is prone to be influenced by the economic benefits in terms of skills and knowledge complementarities driven by immigration, in turn affecting other previously studied outcomes such as economic growth. Since countries' economic complexity increases with the accumulation of a country's stock of capabilities or know-how (as interpreted by Hidalgo and Hausmann (2009) and others), then we can explain that birthplace diversity should be positively related with countries economic complexity *if* new skills available in a more diverse society complement the available ones.

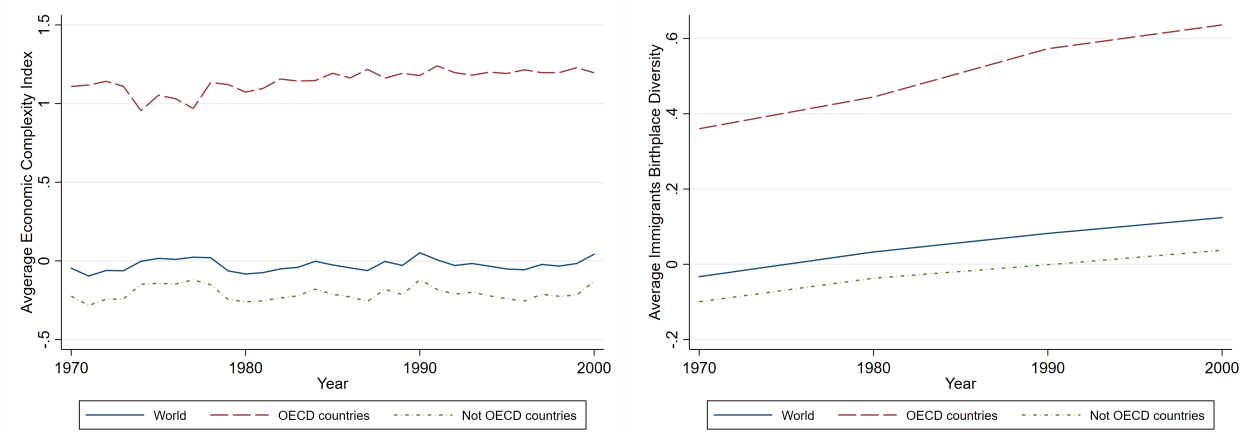
Second, the nature of the ECI allow us to decomposed between export diversification and product uniqueness, allowing us to explore mechanisms through which diversity influences countries' complexity.

Third, even though the ECI has been shown to be an important determinant of income and growth, the evidence so far is thin when it comes to determining which factors influence the economic complexity of countries.

Figure 2.3 plots the evolution of the ECI and immigrants birthplace diversity standardized measures over the period 1970-2000 for the whole World, developed and developing countries.

Figure 2.3(a) shows an overall slight increase of the global average level of economic complexity, although there remain important differences between developed (OECD) and developing (non-OECD) economies. In terms of magnitude, the average level of ECI among developed economies is around ten time larger than the average level of developing economies; while the average yearly growth is larger for developing than for developed economies (1% against 0.4%, respectively). In turn, Figure 2.3(b) plots average birthplace diversity per group of countries. Birthplace diversity follows a positive trend over time, similar to economic complexity, characterized by huge discrepancy between developed and developing economies.

Figure 2.3: Evolution of Economic Complexity and Migrants Birthplace Diversity



(a) Economic Complexity

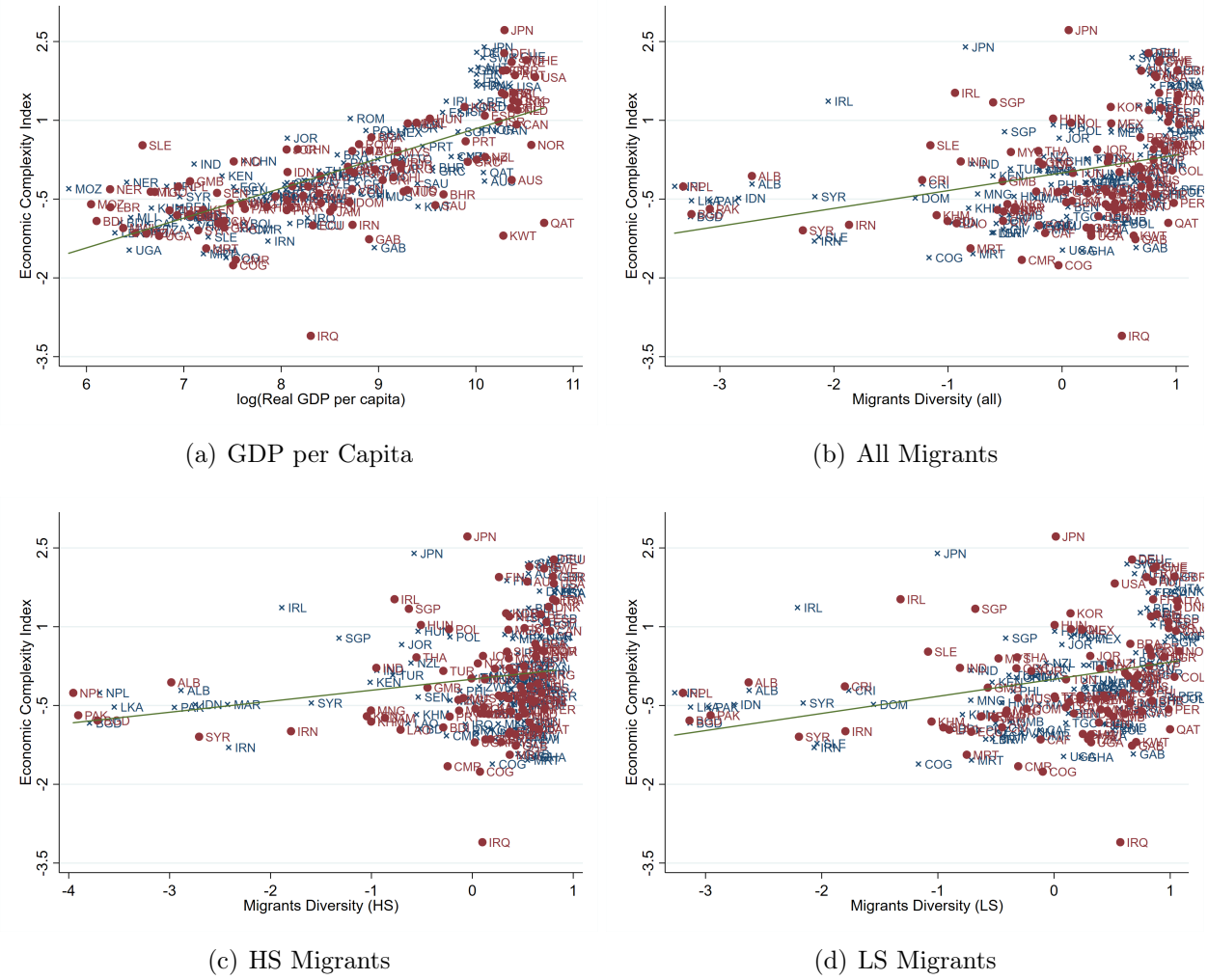
(b) Immigrants Birthplace Diversity

Note: authors' calculations using data from Hausmann *et al.* (2014) and Özden et al. (2011). Figure (a) plots the country average level of the standardized Economic Complexity, while Figure (b) plots the country average level of the standardized Immigrants Birthplace Diversity. Both figures show the World level, OECD countries and Not OECD countries.

Figure 2.4(a) plots the country-year standardized level of ECI over logarithm of GDP per capita, showing a positive and clear relation between the level of complexity and growth across countries, consistently with Hidalgo and Hausmann (2009) and Hausmann *et al.* (2014), among others. Figure 2.4(b) shows a similar although less prominent positive relation between economic complexity and migrants birthplace diversity. This relation remains similar and stable even when we differentiate between highly educated and less educated migrants (Figures 2.4(c) and 2.4(d), respectively).

The descriptive evidence, therefore, points toward a clear positive relation between economic complexity and birthplace diversity. However, a question arise: What is the relation between economic complexity, diversity and economic growth? Knowing from the literature that birthplace diversity positively affects countries' income, as an additional motivational

Figure 2.4: Economic Complexity and Migrants Birthplace Diversity



(a) GDP per Capita

(b) All Migrants

(c) HS Migrants

(d) LS Migrants

Note: authors' calculations using data from Hausmann *et al.* (2014) and Artuc et al. (2015). Figures (a), (b), (c) and (d) plot the country level of the standardized Economic complexity index (y-axis) on the country level of the logarithm of real GDP per capita (fig (a)) and on the standardized migrants birthplace diversity index by skill group. All the correlations are statistically significant. Blue cross shows the country-observation in the 1990, while red dot shows the country-observation in the 2000.

exercise, we replicate in Table 2.1 columns (1), (3) and (5) the main analysis of Alesina et al. (2016), with our sample of countries. Using their same specification and controls, we consistently estimate a strong and positive relation between diversity and GDP per capita across countries (the latter in log terms).[9] In columns (2), (4) and (6) we include as additional

[9]The set of controls includes both time variant and time-invariant controls. The set of time-variant controls includes measures of quality of institutions from Polity IV, population size, human capital and trade openness. Time-invariant controls are measures of absolute latitude, average temperature, landlocked dummies, measures of ethnic, genetic and linguistic diversity and measures of local diseases.

control the ECI of the country. Interestingly the associated coefficient is always positive and significant. Note that when ECI is controlled for there is a significant reduction of the partial correlation of diversity and GDP per capita of around 20%. This suggests that ECI is, in part, one of the channels through which birthplace diversity affects income as shown in Alesina et al. (2016).

In the next section we dig deeper into the relationship between birthplace diversity and economic complexity.

Table 2.1: Replication of Alesina et al. (2016) controlling for Economic Complexity

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable:	OLS	OLS	OLS	OLS	OLS	OLS
$ln(GDP)$	*All Skill*		*High Skill*		*Low Skill*	
Div^{Mig}	0.177***	0.141**	0.133***	0.100*	0.167***	0.135**
	(0.052)	(0.055)	(0.051)	(0.054)	(0.051)	(0.053)
Mig	0.227***	0.214***	0.257***	0.232***	0.221***	0.206***
	(0.067)	(0.062)	(0.064)	(0.066)	(0.068)	(0.063)
ECI		0.196**		0.205**		0.197**
		(0.081)		(0.085)		(0.081)
Observations	200	200	200	200	200	200
Countries	100	100	100	100	100	100
Adj. R-Square	0.90	0.90	0.89	0.89	0.89	0.90
Controls	✓	✓	✓	✓	✓	✓
Regional FE	✓	✓	✓	✓	✓	✓

Note: authors' calculations using data from Hausmann *et al.* (2014) and Alesina et al. (2016). Standard errors are clustered at country level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is the logarithm of real GDP per capita. Each regression includes the full set of controls of Table 3 of Alesina et al. (2016).

2.3 Empirical Strategy

We follow Alesina et al. (2016) as benchmark model for our analysis. We first replace their dependent variable with the $ECI_{d,t}$ as defined above. Our baseline specification is:

$$ECI_{d,t} = \alpha + \beta Div^{Mig}_{d,t,s} + \gamma Mig_{d,t,s} + \delta ln(GDPpc)_{d,t} + \Theta X_{d,t} + \Lambda Z_d + \eta_t + \epsilon_{d,t}. \quad (2.12)$$

Where d indexes a country and t a year. The main variable of interest is $Div^{Mig}_{d,t,s}$, which represents a skill-specific birthplace diversity index among migrants. In particular $s \in \{All, HS, LS\}$, which represent all migrants (All), high-skilled (HS), and low-skilled (LS).

As controls we include a number of variables. First, $Mig_{d,t,s}$, which measures the skill-specific share of migrants in destination country d at year t and skill group s. We also include as an additional control the logarithm of real GDP per capita in the country under consideration (d), which aims to control for the level of development of the countries. We know that, indeed, there is a strong correlation between ECI and income per capita, and thus, by controlling for it, we are estimating whether birthplace diversity can explain higher economic complexity regardless of income per capita. The vector $X_{d,t}$ contains a set of country-time variant characteristics. This exahustive list of controls aims to reduce concerns that our estimates are driven by omitted variable bias from a number of observables. The controls include: (i) quality of institutions measured using the Polity-2 score from the Polity IV database (Marshall *et al.*, 2013), (ii) population size from the UN database (United Nations, 2013) (iii) country-level aggregated years of schooling from Barro and Lee (2013), (iv) real trade openness from the PWT 8.0, (v) trade diversity of imports and exports (measured following Feenstra *et al.*, 2005), and (vi) the weighted average of GDP per capita from all origin countries of migrants to country d using migration shares as weights. The vector Z_d contains a set of country time invariant characteristics, such as: (i) landlocked dummies from Head et al., 2010, (ii) absolute latitude and share of population within 100km of an ice-free coast (iii) average temperature and precipitation, (iv) continent fixed effects, (v) measures of ethnic, linguistic and genetic diversity of the local population from Alesina *e*t al., 2003 and Ashraf and Galor, 2013, and (vi) measures of diseases like malaria, tuberculosis incidence and yellow fever from World Bank, 2013. η_t represents year fixed-effects.

Estimating equation (2.12) using OLS allow us to obtain an estimate of the partial correlation between economic complexity and migration diversity, captured by the estimate of β. However, in spite of the inclusion of the full battery of time-variant and invariant controls for destination countries specified above, it is likely that unobserved factors captured by $\epsilon_{d,t}$ may be correlated both with immigrants' diversity and economic complexity. In that case our estimated OLS coefficient would be a biased estimate of the causal effect of diversity on economic complexity. To mitigate such potential threat we firstly minimize the potential

unobserved factors driven by immigration. To clarify whether the effects of immigration diversity are driven by a diverse set of skills and competences and not by the fact that immigrants are coming from more economically complex societies, we follow the literature on epidemiological effects (Spilimbergo (2009), Docquier et al. (2016) and Valette (2018)) by including as additional control an origin-specific effect $\overline{ECI}^{w}_{d,t,s}$, which is the weighted average of ECI at origin. Using as weights the shares of migrants by skill and origin, $\overline{ECI}^{w}_{d,t,s}$ is a measure of exposure to economic complexity driven by international migration of destination country d at year t. Namely, we compute it as follows:

$$\overline{ECI}^{w}_{d,t,s} = \sum_{o} \Big(\frac{MIG_{d,o,t,s}}{\sum_{o} MIG_{d,o,t,s}} \overline{ECI}_{o}^{1990-2000} \Big) \tag{2.13}$$

where $\overline{ECI}_{o}^{1990-2000}$ is the average ECI at origin country o over the period 1990-2000. If the effect of immigration is mainly driven by the level of economic complexity brought by immigrants, the estimated coefficient related to birthplace diversity should become statistically not significant once that term is controlled for.

Another potential omitted factor related to immigration and diversity is that immigration can have a persistent/long run effect. If that is the case, the partial correlation β estimated in equation (2.12) can include also the effects of previously arrived immigrants. To precisely estimate the effect on economic complexity of recently arrived immigrants, we use Özden et al. (2011)'s data and compute a measure of immigrants birthplace diversity and share in 1960. Since data on the skill composition of immigrants are not available, we compute it over the whole immigrant population. Those controls will capture whether the effect of immigration are driven by previous cohorts of immigrants or by third/second immigration diversity.

Although we minimize the potential bias driven by omitted factors related to immigration, there could still be omitted factors related to destination countries that will bias our estimates. To deal with this issue, we perform two different analysis. First, we include in our cross-country sample country fixed-effects, accounting for time-invariant country specific unobserved heterogeneity. Being aware of the short time dimension of our data set (T=2), the time-invariant factor can capture the majority of the variation associated to the dependent variable, providing imprecise estimates of the coefficients associated to our main variable of interest. Second, to overcome this issue driven by the reduced time dimension, we merge and harmonize the Database on Immigrants in OECD Countries (DIOC) and the Artuc et al. (2015) data to have a short panel of 22 destination countries covering three time periods

(1990, 2000 and 2010).[10] Due to data limitations and the fact that DIOC focuses on developed economies, we cannot cover the same sample as in our main cross-country analysis. However having such panel of countries is quite unique, since for each of them we have the bilateral education specific migration stocks from 194 origin countries,[11] allowing us to build our education-specific birthplace diversity indexes for the relevant immigrant populations. Over this small panel we then estimate the following equation:

$$ECI_{d,t} = \alpha + \beta_p Div^{Mig}_{d,t,s} + \gamma_p Mig_{d,t,s} + \delta_p ln(GDPpc)_{d,t} + \zeta_p HC_{d,t} + \eta_d + \eta_t + \epsilon_{d,t}. \quad (2.14)$$

The estimated coefficient β_p captures the partial correlation of diversity among migrants on economic complexity after controlling for all time-invariant characteristics of destination countries through the inclusion of country fixed-effects η_d. Moreover we also include in the specification a standardized index of human capital, $HC_{d,t}$, based on the average years of schooling attained from Barro and Lee (2013) and taken from the Penn World Table (Feenstra *et al.* (2015)). Including such index allows us to compare the effects of diversity and human capital on economic complexity.[12] Equation (2.14) is empirically demanding, given the small number of countries and the short time dimension of the panel (T=3). However finding significant estimates even in such short panel analysis should minimize concerns related to omitted variable bias.

2.3.1 Identification Strategy

Since immigration is not a random process, another possible threat to our empirical strategy is the presence of a potential reverse causation bias. Countries that are more economically complex are also, on average, richer. Due to better economic conditions, those countries could attract a larger inflow of immigrants coming from a wider range of origin countries. If that is the case, the estimated coefficient β in equation (2.12) will not give us any relevant information about the causal relation between diversity and economic complexity. To deal with reverse causation, we follow the literature and use two different two-stage least square estimators to predict skill-origin specific bilateral stocks of migrants. Our first IV strategy is based on Alesina et al. (2016), who proposed a gravity model to predict the bilateral stocks

[10] The harmonization of those data sets is explained in the 2.8.1

[11] The only exception is Germany, since from DIOC we have information only for 51 origin countries.

[12] Docquier *et al.* (2019b) show that human capital has a bigger role than birthplace diversity in explaining the economic growth of US States over the period 1990-2010.

of immigrants. To minimize the possible violation of the exclusion restriction, the gravity model proposed is quite parsimonious. The model of bilateral migration is defined as:

$$\begin{aligned} MIG_{d,o,t,s} =& \alpha + \gamma_1\ Pop_{d,1960} + \gamma_2\ Dist_{d,o,t} + \gamma_3\ Border_{d,o} + \gamma_4\ Lang\ Off_{d,o} \\ &+ \gamma_5\ Lang\ ethn_{d,o} + \gamma_6\ Colony_{d,o} + \gamma_7\ Time\ Zone_{d,o} + \zeta_{o,t} + \eta_t + \epsilon_{d,o,t,s} \end{aligned} \tag{2.15}$$

where $MIG_{d,o,t,s}$ is the stock of immigrants from country of origin o in destination country d with education s at year t. The set of controls includes the population in destination country d in 1960, the bilateral weighted distance, the presence of a common border, dummies for common official and ethnic minority languages, previous colonial ties and time zone differences.[13] We also include year fixed-effects to capture common trends across countries (η_t) and origin-year specific fixed effects, to capture origin country specific trends ($\zeta_{o,t}$). Due to the high number of zeros given by empty bilateral corridors, we estimate equation (2.15) using a Pseudo-Poisson Maximum Likelihood estimator (PPML) (see Santos Silva and Tenreyro (2006)). We then use the predicted coefficients associated to the estimated gravity equation (2.15) to predict the skill-specific bilateral stocks of migrants ($\widetilde{MIG}^{Grav}_{d,o,t,s}$) and, with them, we compute the predicted skill-specific birthplace diversity indexes ($\widetilde{Div}^{Grav}_{d,t,s}$) and immigration shares ($\widetilde{Mig}^{Grav}_{d,t,s}$). Since the estimated bilateral stocks are less driven by destination countries economic factors, we then use the resulting predicted diversity and share of immigration as IVs.

The second IV approach is based on a shift-share methodology (Card (2001), Ottaviano and Peri (2006), Docquier *et al.* (2019b)). The intuition of this approach is to use past migration settlement patterns as predictor of subsequent migration flows due to network effects. Those predicted flows should be uncorrelated (or at least less correlated) with current levels of economic complexity and development. The variation in aggregate flows of immigrants across origin countries, which is mainly driven by origin-specific push factors, are allocated to our sample of destination countries according to the early distribution of immigrants from the same country of origin. Hence such shift-share instruments produce variation in immigration across destination countries over time due to the interaction between previously established immigrants' settlements and current emigration flows. Following Moriconi *et al.* (2018), we construct skill-specific bilateral stocks taking into account the aggregate variation

[13] Data and methodology from Head et al. (2010) and CEPII. The weighted distance is based on distances between the biggest cities in the countries weighted by the share of the city in the overall country population.

of immigrant flows by skill and origin, and apply it to the same distribution of immigrants by origin. Using Özden et al. (2011) we compute the initial presence of immigrants from origin country o in destination country d in 1970 as share of the total immigrants from the same origin country as follow:

$$sh^{mig}_{d,o,1970} = \frac{MIG_{d,o,1970}}{\sum_d^D MIG_{d,o,1970}} \tag{2.16}$$

where $MIG_{d,o,1970}$ is the stock of migrants from origin country o in destination country d in 1970. Then we compute from the Artuc et al. (2015) data for $t \in \{1990, 2000\}$ the total amount of immigrants coming from country o with education s living in our sample of countries as follow:

$$TOT\ MIG_{o,t,s} = \sum_{d}^{D} MIG_{d,o,t,s}. \tag{2.17}$$

Finally we can compute the predicted bilateral stocks of immigrants in destination country d with education s in year t as follow:

$$\widetilde{MIG}^{SS}_{d,t,s} = TOT\ MIG_{o,t,s} * sh^{mig}_{d,o,1970}. \tag{2.18}$$

We will use the imputed bilateral stocks from equation (2.18) to compute predicted skill-specific measures of diversity ($\widetilde{Div}^{SS}_{d,t,s}$) and migration shares ($\widetilde{Mig}^{SS}_{d,t,s}$), which will be our instrumental variables. Since the initial distribution of immigrants is not skill-specific, this avoids the potential threat of destination countries specific capabilities to attract low/high educated immigrants. However, as pointed out for example by Goldsmith-Pinkham *et al.* (2018) or by Jaeger *et al.* (2018)), persistent local conditions that could influence initial immigrants location and economic complexity of destination countries could threaten our identification. To reduce this possible threat, we also perform our shift-share approach using the initial distribution of immigrants in an earlier period (namely, 1960).

2.4 Main Results

Table 2.2 presents our baseline results from equation (2.12). The reported coefficients capture the relation between diversity and economic complexity. From column (1) to (4) we use as specification the same model as Alesina et al. (2016), while from column (5) to (8) we include also the logarithm of real GDP per capita. Columns (1) and (5) estimates are related to the overall diversity driven by immigrants, while in the other columns we report

education-specific estimates. In column (1) we can appreciate a strong and positive correlation between birthplace diversity and economic complexity. Since the variable of interest and the dependent variable are standardized with mean zero and standard deviation equal to one, we can assess the magnitude of the effects. Given all the other factors equal, a one standard deviation increase in the birthplace diversity index is associated with an increase in the economic complexity index of the destination country by 0.18 standard deviation.

Table 2.2: OLS regression on Economic Complexity Index

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Variable: *ECI*	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Div_{All}^{Mig}	0.179***				0.125*			
	(0.054)				(0.064)			
Mig_{All}	0.068				-0.002			
	(0.075)				(0.071)			
Div_{HS}^{Mig}		0.160***		0.077		0.120**		0.087
		(0.056)		(0.089)		(0.059)		(0.076)
Mig_{HS}		0.119		0.113		0.042		0.081
		(0.083)		(0.111)		(0.089)		(0.112)
Div_{LS}^{Mig}			0.166***	0.101			0.115*	0.045
			(0.053)	(0.082)			(0.062)	(0.084)
Mig_{LS}			0.074	-0.001			0.006	-0.049
			(0.076)	(0.090)			(0.071)	(0.086)
$ln(GDPpc)$					0.306**	0.300**	0.306**	0.301**
					(0.128)	(0.119)	(0.127)	(0.126)
HH Index Trade	0.010	0.135	-0.009	0.069	0.151	0.216	0.141	0.204
	(0.355)	(0.384)	(0.354)	(0.372)	(0.330)	(0.361)	(0.331)	(0.359)
Observations	200	200	200	200	200	200	200	200
Countries	100	100	100	100	100	100	100	100
Adj. R-Square	0.74	0.74	0.73	0.73	0.75	0.75	0.75	0.75
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Regional FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations using data from Hausmann *et al.* (2014) and Alesina et al. (2016). Standard errors are clustered at country level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is a standardized measure of the Economic Complexity Index. Each regression includes the full set of controls of Table 3 of Alesina et al. (2016).

A reasonable prior is that if the main contribution of immigration diversity to countries'

economic complexity is through skill complementarities, then estimates that rely on birthplace diversity among high-skilled migrants should result in a higher point estimate. As Alesina et al. (2016) and Docquier *et al.* (2019b) point out, diversity among college educated immigrants has a positive and significant effect on countries and US states' economic performance while diversity among low-educated immigrants has a smaller and less robust effect. In that sense, investigating the skill-specific relation between diversity and economic complexity provides a more stringent tests of this prior. We present the skill-specific results from columns (2) to (4), and from columns (6) to (8). The relation between economic complexity and skill-specific diversity appears robust across immigrants education groups, and there is no statistical difference between the point estimates of low and high educated diversity indexes. Including both high and low educated birthplace diversity indexes (col. (4)) produces no significant results due to the high correlation between the two measures. However this will no longer be the case (i.e., the two coefficients become statistically different and actually low-skill diversity becomes insignificant) when we conduct a heterogeneity analysis and control for more characteristics in the next set of regressions.

The coefficient for the relative size of the immigrant population (the share of foreign-born) is also positive but not precisely estimated. The inclusion of GDP per capita as an additional control (col. (5) to (8)) results in a reduction of about 25% the size β. However the estimated coefficients remain positive and statistically significant, in particular for diversity among highly educated immigrants.[14] Those results suggest a strong a positive correlation between diversity and economic complexity, regardless of income.[15]

Since our sample includes both developed and developing economies, we can test for non-linearities in the relation between birthplace diversity and economic complexity. Our prior here is that in an economy with a low level of complexity, where the structure of the economy is such that most workers perform manual tasks and work in very small teams or even individually, there may be not much complementarity to expect to start with. Immigrants may well bring new sets of skills and knowledge but these cannot be combined in production due to the structure of the economy. In addition, other complementary inputs such as physical capital, infrastructures or best managerial practices may be missing as well, preventing poor countries from taking advantage of the opportunities linked to diversity.

[14]Note that the inclusion of GDP per capita could be a "bad control" for our analysis if it is determined simultaneously with our index of diversity (see Angrist and Pischke, 2009). Yet, we include it to reduce possible bias due to the established relation between countries' average level of development and their economic complexity.

[15]Since our benchmark analysis includes measures of import and export diversification, our estimates are capturing more than a relation between birthplace diversity and diversification of export and import baskets. Similarly, our estimates are robust to including as an additional control a measure of net inflow foreign net

Table 2.3: OLS regression on Economic Complexity Index - ECI Terciles

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dep. Variable: *ECI*	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
	First Tercile				*Second Tercile*				*Third Tercile*			
Div_{All}^{Mig}	-0.069				0.165**				-0.201			
	(0.140)				(0.071)				(0.147)			
Mig_{All}	0.075				0.065				0.344			
	(0.055)				(0.064)				(0.405)			
Div_{HS}^{Mig}		-0.183		-0.286		0.141***		0.128**		-0.046		0.314
		(0.163)		(0.194)		(0.047)		(0.058)		(0.131)		(0.266)
Mig_{HS}		0.163		0.091		-0.009		-0.072		0.037		0.163
		(0.109)		(0.117)		(0.063)		(0.085)		(0.210)		(0.222)
Div_{LS}^{Mig}			-0.061	0.098			0.142*	0.024			-0.187	-0.462
			(0.142)	(0.095)			(0.078)	(0.087)			(0.128)	(0.287)
Mig_{LS}			0.071	0.078			0.066	0.090			0.253	0.370
			(0.052)	(0.050)			(0.064)	(0.098)			(0.259)	(0.305)
$ln(GDPpc)$	0.012	0.013	0.015	-0.019	-0.030	0.063	-0.013	0.037	1.079***	1.083***	1.065***	0.877***
	(0.110)	(0.113)	(0.111)	(0.119)	(0.115)	(0.074)	(0.118)	(0.125)	(0.214)	(0.202)	(0.219)	(0.213)
Observations	67	67	67	67	67	67	67	67	66	66	66	66
Countries	42	42	42	42	44	44	44	44	36	36	36	36
Adj. R-Square	0.18	0.25	0.18	0.24	0.31	0.37	0.28	0.34	0.80	0.78	0.80	0.80
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Regional FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations using data from Hausmann *et al.* (2014) and Alesina et al. (2016). Standard errors are clustered at country level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is a standardized measure of the Economic Complexity Index. Each regression includes the full set of controls of Table 3 of Alesina et al. (2016) plus the logarithm of GDP. The sample of countries is splitted by ECI terciles: first tercile (col. (1)-(4)), second tercile (col (5)-(8)) and third tercile (col. (9)-(12)).

In an attempt to capture potential heterogeneous effects, we therefore split our sample of countries by terciles of ECI and perform a subsample analysis for each tercile of the distribution. Table 2.3 presents the results by tercile. Columns (1) to (4) present the estimates for the first tercile, columns (5) to (8) for the second tercile and finally columns (9) to (12) for the upper tercile of the distribution. As in Table 2.2, we present the estimates both for overall immigration diversity and for education-specific immigrants birthplace diversity. There are two main results we can take from Table 2.3. First, immigrants birthplace diversity, in particular among highly educated immigrants, is positively and strongly associated with economic complexity of countries at intermediate levels of economic complexity (namely, in the 2^{nd} tercile of the distribution). Birthplace diversity among highly educated immigrants is strongly significant also in the more demanding specification where we control for both skill-specific measure of diversity (col. (8)). The correlation between diversity and economic complexity is not significant in the other terciles of the distribution. Those results suggest

investment. These results are available upon request.

that diversity contributes to the accumulation of knowledge and abilities required to build up the economic complexity of developing countries but is less important for developed or underdeveloped countries. Second, real GDP per capita is highly correlated with economic complexity only in the upper tercile of the distribution. Intuitively, this result suggests that birthplace diversity is a factor explaining economic complexity but only when the local population has the *right* amount of skills and competences to be complemented by those brought by the immigrants.

We further test whether the positive and significant relation between diversity among immigrants and economic complexity is robust to using alternative measures of diversity. Table A2 in 2.8.2 shows the estimates over the sample of countries belonging to the 2^{nd} ECI tercile after including the number of immigrants' countries of origin represented (col. (2)), and after replacing the diversity indexes with polarization indexes (col. (3)) and Theil indexes and its decomposition (col. (4) to (6)). The estimates confirms the direction of the estimates: lower concentration (as measured by the Theil index , which is consistent with high diversity) is associated with higher economic complexity. Moreover, the significant coefficient associated to the between component of the Theil index (col. (6)) suggests that the positive relation between diversity and economic complexity is mainly driven by an expansion of the set of origin countries, which implies an introduction of new skills and competences in the destination country.

As explained in Section 2.3, the strong partial correlation found between diversity and complexity could be driven by unobserved factors. Immigrants coming from a wide range of origin countries can bring with them a wide set of skills and competences, but they can also bring with them knowledge driven by the economic complexity of their origin country. To clarify whether the effect of immigration on economic complexity is driven by diversity, following Valette (2018) we include in our specification a measure of exposure to economic complexity driven by international migration ($\overline{ECI}_s^w$). Table 2.4 presents the estimates for the full sample (col. (1) to (6)) and for the subsample of countries in the 2^{nd} tercile of ECI (col. (7) to (12)). Replacing the skill-specific birthplace diversity index with the skill-origin-specific term does not produce any statistically significant relation between the latter term and economic complexity, as odd columns in Table 2.4 show. Once we include simultaneously the birthplace diversity index and the origin-specific term, the partial correlation between birthplace diversity and economic complexity remains positive and statistically significant. The correlation is stronger among countries belonging to the 2^{nd} tercile of the ECI distribution. Moreover the origin-specific term is not statistically significant in the full sample, while it is negatively correlated with economic complexity in the second tercile subsample. Those

results suggest that the effect of immigration on economic complexity is driven by the variety of complementary skills and competences captured by birthplace diversity, and not by a simple transfer of complexity from origin to destination countries. Moreover, the negative and significant estimates associated to the origin-specific term among countries belonging to the 2^{nd} ECI tercile suggests that the beneficial relation between migration and complexity is not necessarily driven by an increase of skills and competences from complex societies, but from less complex counties. However, such negative relation can be also explained by the fact that those countries are attracting more immigrants from low economically complex countries. Investigating the effect of diversity in a panel of the 51 US states, Docquier *et al.* (2019b) find a similar negative relation between the origin-specific term and states economic growth.

Table 2.4: OLS regression on Economic Complexity Index
Origin Specific Effect

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dep. Variable: *ECI*	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Sample:	*Overall Sample*						2^{nd} *Tercile*					
Education	*All*	*All*	*HS*	*HS*	*LS*	*LS*	*All*	*All*	*HS*	*HS*	*LS*	*LS*
Div_s^{Mig}		0.128**		0.123**		0.117*		0.162***		0.157***		0.148***
		(0.064)		(0.060)		(0.061)		(0.052)		(0.045)		(0.050)
Mig_s	-0.018	-0.005	0.026	0.033	-0.008	0.003	0.022	0.033	-0.026	-0.078	0.019	0.034
	(0.075)	(0.071)	(0.094)	(0.093)	(0.076)	(0.072)	(0.068)	(0.056)	(0.070)	(0.060)	(0.067)	(0.057)
$\overline{ECI}_s^w$	0.007	-0.009	-0.002	0.006	0.028	0.015	-0.116	-0.257***	-0.147	-0.233**	-0.121	-0.256***
	(0.111)	(0.110)	(0.112)	(0.109)	(0.113)	(0.114)	(0.086)	(0.078)	(0.127)	(0.088)	(0.076)	(0.079)
Observations	200	200	200	200	200	200	67	67	67	67	67	67
Countries	100	100	100	100	100	100	44	44	44	44	44	44
Adj. R-Square	0.74	0.75	0.74	0.75	0.74	0.75	0.24	0.44	0.24	0.51	0.25	0.43
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Regional FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations using data from Hausmann *et al.* (2014) and Alesina et al. (2016). Standard errors are clustered at country level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is a standardized measure of the Economic Complexity Index. Each regression includes the full set of controls of Table 3 of Alesina et al. (2016), excluding the weighted average of GDP per capita at origin countries using migration shares as weights. The skill-specific origin-specific term, computed as the weighted average of the Economic Complexity at the origin, where the weights are the migration shares by origin country (see equation (2.13)) is included in all the regressions.

Another potential threat that could bias our estimates is the long-run/persistent effect of immigration. As pointed out by Alesina et al. (2016), our measure of birthplace diversity could also capture the effect of immigrants previously arrived in the destination country, and so our estimates would capture a combination of diversity driven by recent and older immigration. We thus include as additional control a measure of lagged diversity, to capture the effect of diversity driven by previous cohorts. We use Özden et al. (2011) data and construct as additional controls measures of immigrants diversity and shares in 1960. Table 2.5 presents the estimates on the full sample (col. (1)-(4)) and on the 2^{nd} tercile (col. (5)-

(8)). Including past diversity instead of current diversity in columns (1) and (5) shows that past diversity is not significantly correlated with the current level of economic complexity. The inclusion of past diversity as additional control in columns (2) to (4) affects the precision of the estimates related to current diversity in the full sample. This result suggests that part of the strong correlation between current diversity and economic complexity is driven by previous cohorts of immigrants. Not surprisingly, current and past diversity are strongly correlated (at 0.70). However, the estimates for current immigration diversity over the subsample of countries belonging to the 2^{nd} tercile are positive and significant, in particular among highly educated immigrants (col. (6)-(8)). This result confirms the relevant role of diversity in skills and competences driven by immigration in developing economies, expressed by a strong a positive partial correlation between diversity and economic complexity.

Table 2.5: OLS regression on Economic Complexity Index
Robustness by previous level of diversity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Variable:	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
ECI								
Sample:	*Overall Sample*				2^{nd} *Tercile*			
Education:	*All*	*All*	*HS*	*LS*	*All*	*All*	*HS*	*LS*
$Div^{Mig}_{All,1960}$	0.107	0.013	0.044	0.027	0.116	-0.025	0.003	0.020
	(0.079)	(0.098)	(0.088)	(0.099)	(0.073)	(0.089)	(0.078)	(0.090)
$Mig_{All,1960}$	0.024	0.044	0.015	0.034	0.063	0.013	0.043	0.017
	(0.076)	(0.088)	(0.080)	(0.089)	(0.073)	(0.086)	(0.062)	(0.092)
Div^{Mig}_{s}		0.116	0.099	0.097		0.182*	0.142**	0.131
		(0.079)	(0.067)	(0.077)		(0.097)	(0.056)	(0.102)
Mig_{s}		-0.032	0.034	-0.017		0.060	-0.019	0.059
		(0.073)	(0.090)	(0.073)		(0.076)	(0.067)	(0.078)
$ln(GDPpc)$	0.336***	0.307**	0.296**	0.308**	0.079	-0.032	0.056	-0.014
	(0.116)	(0.127)	(0.121)	(0.127)	(0.087)	(0.118)	(0.097)	(0.123)
Observations	200	200	200	200	67	67	67	67
Countries	100	100	100	100	44	44	44	44
Adj. R-Square	0.75	0.75	0.75	0.75	0.21	0.27	0.34	0.24
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Regional FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations using data from Hausmann *et al.* (2014) and Alesina et al. (2016). Standard errors are clustered at country level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is a standardized measure of the Economic Complexity Index. Each regression includes the full set of controls of Table 3 of Alesina et al. (2016).

2.5 Robustness

After assessing a stable and positive correlation between immigrants diversity, particularly among highly educated immigrants, and economic complexity, this section investigates the robustness of those results.

We first test the robustness of our estimates to time-invariant unobserved heterogeneity, by including in our sample country fixed-effects. The inclusion of time-invariant factors with

a reduced time dimension (T=2) is extremely demanding from an empirical point of view, since the majority of the variation can be captured by country fixed-effects. Being aware of that, Table 2.6 presents the skill-specific estimates over the whole sample (col. (1) to (4)) and over the sample of countries belonging to the 2^{nd} tercile of the Economic Complexity index. Even though the estimates on the overall sample are not precisely estimated, columns (6) and (8) show that the positive relation between diversity among college educated immigrants and economic complexity is robust to the inclusion of time-invariant country-specific factors. However, given the small sample size and the short time dimension, those estimates should be interpreted cautiously.

Table 2.7 presents the results of our short panel over OECD developed countries. Due to data limitations, our short panel cover 22 destination countries[16] over three periods (1990, 2000 and 2010). However, thanks to the bilateral structure and the skill composition available in the harmonized data set we are able to compute skill specific measures of birthplace diversity for each destination country. Odd columns present the estimates of equation (2.14) without country specific control, while even columns include as additional control measures of human capital and GDP per capita. Moreover all the estimates include year and country fixed effects, controlling for common trend and for time-invariant country specific factors. The main results of this robustness are two. First, there is a strong and positive correlation between birthplace diversity and economic complexity also in a panel dimension. The results are stronger among highly educated immigrants. Including country specific controls enhances the effect of diversity on economic complexity. Those results should minimize the possible concerns related to an omitted variable bias. Moreover, the negative and statistically significant coefficient associated to the share of immigrants can be explained by countries characterized by a sizeable growth of economic complexity while almost absent or negative growth in terms of immigration share, due to restrictive migration policies (e.g. Japan and Mexico). Second, human capital is another strong and good predictor of economic complexity. Since the measure of human capital is standardized as birthplace diversity,[17] we can compare the magnitude of the two coefficients. Indeed, the effect of human capital on economic complexity are on average 3 times bigger than the effect of birthplace diversity. Those results are in line with Docquier *et al.* (2019b), which finds, using US data, that the effect of human capital on GDP growth is up to four times larger than birthplace diversity measured on the college educated immigrants.

[16]The 22 destination countries in analysis are: Australia, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Japan, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Spain, Sweden, United Kingdom, United States.

[17]Both measures are standardized with mean zero and standard deviation equal to one.

Table 2.6: OLS regression on Economic Complexity Index - Country fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Variable:	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
ECI								
Sample:	*Overall Sample*				2^{nd} *Tercile*			
Div_{All}^{Mig}	0.105				-0.026			
	(0.084)				(0.195)			
Mig_{All}	-0.348*				-0.039			
	(0.206)				(0.451)			
Div_{HS}^{Mig}		0.022		-0.087		0.198**		0.397***
		(0.052)		(0.091)		(0.081)		(0.138)
Mig_{HS}		-0.017		-0.040		-0.483***		-0.658***
		(0.119)		(0.113)		(0.166)		(0.215)
Div_{LS}^{Mig}			0.106	0.175			-0.024	-0.371*
			(0.077)	(0.126)			(0.200)	(0.192)
Mig_{LS}			-0.212	-0.207			0.067	-0.073
			(0.156)	(0.159)			(0.391)	(0.219)
$ln(GDPpc)$	-0.005	-0.090	-0.054	-0.024	0.007	0.041	-0.021	0.117
	(0.192)	(0.220)	(0.193)	(0.186)	(0.223)	(0.131)	(0.223)	(0.125)
Observations	200	200	200	200	46	46	46	46
Countries	100	100	100	100	23	23	23	23
Adj. R-Square	0.93	0.92	0.92	0.92	0.42	0.60	0.42	0.61
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations using data from Hausmann *et al.* (2014) and Alesina et al. (2016). Standard errors are clustered at country level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is a standardized measure of the Economic Complexity Index. Each regression includes the full set of controls of Table 3 of Alesina et al. (2016) plus the logarithm of GDP. Moreover, each specification includes country fixed effects.

In Table 2.8 we test whether our results hold after instrumenting our main variable of interest. As it is presented in Section 2.3, we have two different sets of instrumental variables. Descriptive statistics across different standardized instrumental variables and actual birthplace diversity and immigration share are available in Table A4, Table A5 and Table A6 in 2.8.2. Overall there is a strong correlation across different measures of diversity (actual and predicted). However, looking at correlations in Tables A5 and A6, we can see that both IV

Table 2.7: Panel Analysis on Economic Complexity Index

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Variable: ECI	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Div_{All}^{Mig}	0.120**	0.159***						
	(0.045)	(0.043)						
Mig_{All}	-0.353**	-0.300**						
	(0.162)	(0.130)						
Div_{HS}^{Mig}			0.085**	0.151**			0.197	0.286*
			(0.041)	(0.066)			(0.145)	(0.155)
Mig_{HS}			-0.132	-0.033			-0.157*	-0.057
			(0.081)	(0.095)			(0.083)	(0.092)
Div_{LS}^{Mig}					0.052	0.104**	-0.076	-0.102
					(0.048)	(0.049)	(0.117)	(0.127)
Mig_{LS}					-0.097	-0.104	-0.153	-0.145
					(0.082)	(0.095)	(0.095)	(0.101)
HC		0.517**		0.559**		0.563**		0.554**
		(0.201)		(0.246)		(0.214)		(0.236)
$ln(GDPpc)$		-0.008		-0.317		-0.033		-0.323
		(0.307)		(0.545)		(0.414)		(0.449)
Observations	66	66	66	66	66	66	66	66
Countries	22	22	22	22	22	22	22	22
Adj. R-Square	0.95	0.96	0.95	0.96	0.95	0.96	0.95	0.96
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations using data from Hausmann *et al.* (2014), Artuc et al. (2015), DIOC and Penn World Table. Standard errors are clustered at country level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is a standardized measure of the Economic Complexity Index. Each regression includes country and year fixed effects. HC is the standardized measure of human capital (Penn World Table). $ln(GDP)$ is the logarithm of real GDP per capita. The list of countries in the sample is the following: Australia, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Japan, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Spain, Sweden, United Kingdom, United States.

approaches have some issues with the prediction of the stocks of highly educated immigrants. The gravity model is not able to predict properly the composition of the bilateral stocks of highly educated migrants in terms of countries of origin (i.e. diversity index), but it is able to produce a good predictor of the skill-specific share of immigrants. On the other hand,

Table 2.8: IV regression on ECI

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Variable: *ECI*									
Panel A - Overall Sample									
Div_{All}^{Mig}	0.221*	0.116**	0.219*						
	(0.113)	(0.059)	(0.114)						
Mig_{All}	0.003	-0.014	-0.003						
	(0.064)	(0.065)	(0.064)						
Div_{HS}^{Mig}				0.260**	0.110**	0.261**			
				(0.115)	(0.054)	(0.119)			
Mig_{HS}				0.045	-0.064	-0.038			
				(0.086)	(0.135)	(0.135)			
Div_{LS}^{Mig}							0.189*	0.109*	0.195*
							(0.108)	(0.057)	(0.107)
Mig_{LS}							0.008	0.020	0.029
							(0.064)	(0.073)	(0.071)
Observations	200	200	200	200	200	200	200	200	200
Countries	100	100	100	100	100	100	100	100	100
K-P rk Wald F-stat	13.97	425.25	7.07	16.77	29.41	8.79	16.49	126.61	8.34
Adj. R-Square	0.74	0.75	0.74	0.74	0.75	0.73	0.74	0.75	0.74
Panel B - 2^{nd} Tercile									
Div_{All}^{Mig}	0.223***	0.176***	0.228***						
	(0.053)	(0.048)	(0.054)						
Mig_{All}	0.091*	0.096*	0.106*						
	(0.051)	(0.051)	(0.055)						
Div_{HS}^{Mig}				0.166***	0.147***	0.159***			
				(0.047)	(0.036)	(0.050)			
Mig_{HS}				-0.002	-0.066	-0.066			
				(0.046)	(0.058)	(0.056)			
Div_{LS}^{Mig}							0.224***	0.159***	0.235***
							(0.058)	(0.052)	(0.058)
Mig_{LS}							0.101*	0.112**	0.132**
							(0.056)	(0.056)	(0.066)
Observations	67	67	67	67	67	67	67	67	67
Countries	44	44	44	44	44	44	44	44	44
K-P rk Wald F-stat	40.47	517.43	17.69	25.91	15.95	10.73	24.48	626.49	10.89
Adj. R-Square	0.29	0.30	0.29	0.35	0.33	0.32	0.24	0.27	0.23
Instr *Div*	✓		✓	✓		✓	✓		✓
Instr *Mig*		✓	✓		✓	✓		✓	✓

Note: authors' calculations using data from Hausmann *et al.* (2014), Alesina et al. (2016) Özden et al. (2011), and Artuc et al. (2015). Standard errors are clustered at country level. * p<0.1, ** p<0.05, *** p<0.01. This table shows the effect of birthplace diversity on our standardized measure of Economic Complexity,after controlling for the full set of controls of Table 3 of Alesina et al. (2016) and for the logarithm of real GDP per capita. Birthplace diversity is instrumented with predicted stocks through shift-share methodology based on the 1970 distribution, while the migration share is instrumented with predicted stocks from gravity model.

the shift-share approach is able to predict the composition in terms of countries of origin of

college-educated immigrants, while it fails to predict the size of immigrants population.[18] To properly identify both endogenous variables we then instrument diversity indexes with IVs based on predicted stocks from the shift-share approach, while we instrument skill-specific shares of immigrants with IVs based on predicted stocks from the gravity model. Table 2.8 reports the second-stage estimates of our IV approach. Columns (1), (4) and (7) show the estimates when we instrument only the skill-specific diversity index, while columns (2), (5) and (8) present the estimates when the skill-specific share of immigrants is treated as endogenous variable. Finally, columns (3), (6) and (9) shows the estimates when both variables are simultaneously instrumented. Moreover our analysis is done over the whole sample of countries (Panel A) and the subsample of countries belonging to the 2^{nd} tercile of the economic complexity distribution (Panel B). Looking at both panels, the instrumental variables have enough predictive power to explain both endogenous variables when we instrument them separately or simultaneously. The F-stat are always above the critical values specified by Stock and Yogo (2002). The estimates associated to birthplace diversity indexes are always positive and significant on the overall sample (Panel A), confirming the results presented in Table 2.2. The estimates presented in Panel B are positive and precisely estimated at 1% level. Moreover, all the 2SLS point estimates are larger than the OLS estimates presented in Table 2.3 when we instrument both endogenous variables, suggesting the correction of an attenuation bias.

Overall we can see that the coefficients associated to birthplace diversity are positive and significant across different regressions, confirming the strong and positive relation between diversity and economic complexity in countries that belongs to the middle of the economic complexity distribution.[19]

2.6 Discussion

Our analysis shows a clear and strong relation between immigrants birthplace diversity and economic complexity. Our estimates are stronger when focusing on college educated immigrants and among developing countries. Those results are robust to the inclusion of historical diversity, origin-specific effects, country fixed-effects and panel analysis. Combining those robustness with our IV estimates, we mitigate the potential reverse causality and omitted variable bias threats, reducing concerns of our results being driven by endogeneity.

[18] Docquier *et al.* (2019b) recognize the same weakness of the shift-share methodology to predict the share of college-graduate immigrants.

[19] Second-stage estimates with alternative shift-share stocks based on the 1960 distribution are available in Table A7 in 2.8.2.

However, some questions remain open. For instance, what are the mechanisms through which immigrants diversity positively affects economic complexity? Empirically determining these mechanisms is an important part of our future agenda, but here we provide some insights.

Simply put, a higher Economic Complexity Index is driven by two components. First, a more diverse export basket in terms of the number of industries that are exported competitively; and second, a smaller average number of countries that export those industries competitively (a low ubiquity of those industries). We extend our analysis by testing whether birthplace diversity of a country contributes both to its export basket diversity and the uniqueness of its products.

Following the literature, we compute different indices to measure the diversification of a country's export basket (see Imbs and Wacziarg (2003), Cadot *et al.* (2011), Bahar and Santos (2018)) using -consistently with the construction of ECI above- 4-digit classification of industries according to SITC. We first construct an index that measures concentration of a country's export basket (e.g., a larger value implies higher concentration, or less diversification), namely the Herfindahl-Hirschman index (HHI here onwards).[20] Assuming there are $i = 1, 2, ..., n$ products and $d = 1, 2, ..., m$ countries, the HHI for any country d in time t is computed using the following formula:

$$HHI_{d,t} = \sum_{i=1}^{n} s_{i,d,t}^2 \tag{2.19}$$

where $s_{i,d,t}$ is the share product i in country d's export basket at time t. We complement it with two more measures. First, the total number of 4-digit SITC industries exported with any value higher than zero; as well as with the total number of products exported by a country d at time t with RCA values above one, consistent with the economic complexity literature. Naturally, for these two measures a higher value implies more export diversification (and less concentration). These measures can be mathematically expressed as:

[20] We also compute alternative indices of product diversification: the Gini coefficient, the Theil Index with its decomposition in the Theil-within and Theil-between. Changes in the Theil-within can be interpreted as changes in concentration due to more concentration of industries that already existed in the export basket of the country, while the Theil-between can be interpreted as changes in export concentration due to the appearance or disappearance of export lines. Results related to those indices and others are reported at Tables A8 and A10 in the 2.8.2.

$$products_{d,t} = \sum_{i=1}^{n} \times 1[exp_{i,d,t} > 0] \tag{2.20}$$

$$productsRCA_{d,t} = \sum_{i=1}^{n} \times 1[rca_{i,d,t} \geq 1] \tag{2.21}$$

We then also compute a number of indices to measure average ubiquity of countries' exports. Firstly we compute the concentration of each product across global export basket, applying the HHI index formula to a product p, as follows:

$$HHI_{p,t} = \sum_{j=1}^{m} s_{p,j,t}^2 \tag{2.22}$$

where j represents a country, and $s_{p,j,t}$ is the share of product p in country j's export basket at time t. A higher value of $HHI_{p,t}$ implies that exports of product p in time t are concentrated on fewer countries. We then compute a country level measure which we name HHI^p, and is constructed as:

$$HHI_{d,t}^{p} = \sum_{i=1}^{n} s_{i,d,t} \times HHI_{i,t} \tag{2.23}$$

In other words, $HHI_{d,t}^{p}$ is a sum of *product-level* concentration indices weighted by the relative size of each product in country's d export basket at time t. A higher value of $HHI_{d,t}^{p}$ implies that country d exports goods that are exported by fewer countries at time t. Consistently with above, we produce similar measures for number of countries that export the product p based on the following two definitions:

$$countries_{p,t} = \sum_{j=1}^{m} \times 1[exp_{p,j,t} > 0] \tag{2.24}$$

$$countriesRCA_{p,t} = \sum_{j=1}^{m} \times 1[rca_{p,j,t} \geq 1] \tag{2.25}$$

Then we use these measures to construct a country-level variable that computes the average export shares weighted by number of countries that export each product exported by country c as follow:

$$products^{p}_{d,t} = \sum_{i=1}^{n} s_{i,d,t} \times countries_{i,t} \tag{2.26}$$

$$productsRCA^{p}_{d,t} = \sum_{i=1}^{n} s_{i,d,t} \times countriesRCA_{i,t} \tag{2.27}$$

In other words, *lower* levels of $products^{p}_{d,t}$ and of $productsRCA^{p}_{d,t}$ imply that country-year pair d and t is exporting products that, on average, are less ubiquitous.

Table 2.9: Correlations of diversity, diversification, ubiquity, income and complexity

	(1) ECI	(2) $ln(GDPpc)$	(3) Div^{Mig}_{All}	(4) Div^{Mig}_{HS}	(5) Div^{Mig}_{LS}	(6) HHI	(7) $products$	(8) $productsRCA$	(9) HHI^{p}	(10) $products^{p}$
ECI	1									
$ln(GDPpc)$	0.720***	1								
Div^{Mig}_{All}	0.338***	0.447***	1							
Div^{Mig}_{HS}	0.209**	0.295***	0.869***	1						
Div^{Mig}_{LS}	0.326***	0.427***	0.993***	0.841***	1					
HHI	-0.557***	-0.308***	-0.109	-0.0477	-0.0872	1				
$products$	0.697***	0.716***	0.342***	0.144*	0.321***	-0.568***	1			
$productsRCA$	0.811***	0.584***	0.305***	0.171*	0.291***	-0.613***	0.754***	1		
HHI^{p}	0.137	-0.0717	0.00592	0.0660	0.00936	-0.353***	-0.0348	0.201**	1	
$products^{p}$	0.306***	0.139*	0.00228	-0.0268	-0.0221	-0.339***	0.316***	0.369***	-0.213**	1
$products\ RCA^{p}$	-0.499***	-0.364***	-0.188**	-0.175*	-0.185**	0.134	-0.290***	-0.458***	-0.402***	0.244***

Note: authors' calculations using data from Hausmann *et al.* (2014), Alesina et al. (2016), and Artuc et al. (2015). This table show correlations among variables with their level of significance * p<0.1, ** p<0.05, *** p<0.01.

Table 2.9 shows the correlation across those measures of diversification, export uniqueness and our measures of birthplace diversity, GDP per capita as well as ECI. The table documents that, naturally, all diversification measures are highly and significantly correlated among each other. But, in addition, the table also documents that economic complexity and immigration diversity are positively correlated with measures of export diversification (or, in fact, negatively correlated with measures of *concentration*). Concerning measures of export uniqueness we find a negative and significant correlation between our weighted measures of competitive products and diversity, suggesting that more diversity is related with less ubiquitous exported goods.

Table 2.10 estimates equation (2.12), but this time uses the export basket diversification and uniqueness measures as dependent variable. The top panel presents the results for the diversification measures, while the bottom panel presents the estimates for the ubiquity measures. Panel A shows across most specifications that birthplace diversity is associated with higher diversification (again, note that point estimates are negative when the dependent variable is a concentration index and positive when using number of products). While

not all the point estimates are statistically different from zero based on the conventional confidence intervals, we do find significant coefficients for the number of products with RCA above one (columns 7 to 9).[21] For instance, an increase of one standard deviation in our birthplace diversity index is associated with an increase in the number of industries exported competitively by around 12 (roughly a 10 percent increase, based on the sample mean of about 120, with standard deviation equal to 81).[22] Panel B shows the partial correlations between skill-specific migration diversity and export ubiquity. Overall the estimates are not statistically different from zero, however the direction of the point estimates is toward a reduction of products ubiquity.[23]

All in all, our evidence points to birthplace diversity explaining higher levels of complexity, in part, through the diversification of country's export basket, though the evidence is not very robust across diversification measures.

[21] Removing from the set of controls measures of diversity in import and export does not affect the size and the precision of the estimates significantly.

[22] Table A10 presents the results for alternative diversification measures. The direction of the results is similar, and we find significant coefficients associated with the Gini index. From the estimates, an increase of one standard deviation of birthplace diversity index is associated with a drop of the Gini coefficient around 0.012 (roughly 1.3 percent based on the sample mean equal to 0.93).

[23] Table A9 proposes an alternative way to test the effect of diversity on products ubiquity. Using countries averages of Product Complexity Index over the top 3, 5 and 10 products in terms of exports, we test whether the best products export become more unique. The Product Complexity Index –an index that mirrors the ECI– measures how much a given product is complex: higher values implies products that are exported by fewer countries and that those countries are highly diversified. Even though the direction of the estimates indicates higher uniqueness, the estimates are not statistically different from zero.

Table 2.10: OLS regression on Diversification and Ubiquity Measures

	HHI			$products$			$productsRCA$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A - Diversification									
Div_{All}^{Mig}	-0.007			7.975			11.438**		
	(0.019)			(11.653)			(4.558)		
Mig_{All}	-0.024			-5.090			1.525		
	(0.020)			(14.611)			(4.615)		
Div_{HS}^{Mig}		-0.006			-11.402			12.687***	
		(0.018)			(12.545)			(4.567)	
Mig_{HS}		-0.033			-7.664			-0.544	
		(0.031)			(15.428)			(6.356)	
Div_{LS}^{Mig}			-0.006			8.248			10.354**
			(0.018)			(11.028)			(4.464)
Mig_{LS}			-0.022			-5.235			2.488
			(0.019)			(14.274)			(4.682)
$ln(GDPpc)$	0.040	0.040	0.038	42.394*	52.321**	42.326*	8.637	10.891	8.543
	(0.026)	(0.029)	(0.026)	(21.602)	(20.421)	(21.372)	(7.281)	(7.248)	(7.322)
Observations	200	200	200	200	200	200	200	200	200
Countries	100	100	100	100	100	100	100	100	100
Adj. R-Square	0.43	0.44	0.43	0.79	0.79	0.79	0.79	0.79	0.79
	HHI^p			$products^p$			$productsRCA^p$		
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Panel B - Ubiquity									
Div_{All}^{Mig}	0.000			0.306			-0.025		
	(0.002)			(1.533)			(0.801)		
Mig_{All}	0.008***			-2.519*			-2.171***		
	(0.002)			(1.358)			(0.693)		
Div_{HS}^{Mig}		0.002			1.567			-0.734	
		(0.002)			(1.418)			(0.757)	
Mig_{HS}		0.008***			-2.017			-2.451**	
		(0.002)			(2.029)			(0.942)	
Div_{LS}^{Mig}			0.000			-0.063			0.020
			(0.002)			(1.512)			(0.798)
Mig_{LS}			0.008***			-2.424*			-2.116***
			(0.001)			(1.341)			(0.663)
$ln(GDPpc)$	-0.014***	-0.013***	-0.013***	1.801	0.737	1.953	-1.398	-1.217	-1.449
	(0.004)	(0.003)	(0.004)	(2.449)	(2.478)	(2.444)	(1.586)	(1.525)	(1.578)
Observations	200	200	200	200	200	200	200	200	200
Countries	100	100	100	100	100	100	100	100	100
Adj. R-Square	0.45	0.43	0.44	0.78	0.78	0.78	0.59	0.59	0.59
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Regional FE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations using data from Hausmann *et al.* (2014) and Alesina et al. (2016). Standard errors are clustered at country level. * p<0.1, ** p<0.05, *** p<0.01. This table shows the effect of birthplace diversity on different measures of product diversification and ubiquity: Herfindahl-Hirschman Index on products exported (col. (1)-(3)), number of exported products (col. (4)-(6)), number of exported products with RCA>1 (col. (7)-(9)), weighted measures of countries product concentration (col. (10)-(12)), weighted sum of products uniqueness (col. (13)-(15)) and weighted sum of products with RCA bigger than one uniqueness (col. (16)-(18)).

2.7 Conclusion

We investigate the relationship between immigration diversity and economic complexity with the prior that a more diverse immigration, especially if immigrants are highly-skilled, brings with it new knowledge and skills that can serve to expand the economic complexity of the receiving economy (i.e., its capacity to develop a set of competitive and unique industries).

We address this question in a sample of 100 countries over the 1990-2000 period for which we could collect data on their levels of economic complexity and on the structure (by country of origin and skill level) of their immigration. In line with the literature, we use the Herfindahl index of diversity in our baseline regressions; we also use the Theil index and its decomposition to refine our interpretation of the results in the robustness section.

We find that the birthplace diversity of immigrants is strongly and positively associated with countries' economic complexity. Increasing birthplace diversity by one standard deviation is associated with an increase in economic complexity by 0.18 standard deviations. This holds particularly true among college-educated migrants and for countries at intermediate levels of economic complexity. Moreover, the results hold in a short panel of developed countries, accounting for countries time-invariant unobserved factors, suggesting that a minimum threshold of complexity should be reached to take advantage of the opportunities linked to diversity.

Those results are robust to controlling for past diversity, country-fixed effects, and for origin-specific effects (i.e., for whether immigrants come from countries which are themselves complex or not). They are also robust to instrumenting migration using two different IV strategies, one based on a pseudo-gravity model, and one of the shift-share methodology. When we use the Theil index, we find that the positive relationship between diversity and complexity is mostly driven by the "between" component of the index, suggesting that the extensive margin of immigration (that is, the diversity of origins) matters. This is consistent with our interpretation of the results in terms of skill complementarity. Finally, we show that the results are driven by the diversification component of the economic complexity index rather than by its uniqueness component Considering the full set of results, we conclude that immigration diversity is an important building block of economic complexity.

2.8 Appendix

2.8.1 Harmonization of DIOC and ADOP data

To minimize the potential bias due to omitted variable, we combined data of DIOC (Database on Immigrants in OECD counties) (`http://www.oecd.org/els/mig/dioc.htm`) and Artuc et al. (2015) data (`https://perso.uclouvain.be/frederic.docquier/filePDF/ADOP_AgeOfEntry.xlsx`), to have a panel of 22 destination countries over the period 1990 to 2010. Since DIOC is available over 2000 and 2010 while ADOP is available over 1990 and 2000, we had to harmonize our data sources. Moreover we take DIOC data as main reference in order to harmonize ADOP 1990 data with the rest of DIOC data. Our harmonization process took the following steps:

1. we compute the bilateral skill specific weight in the year 2000, such that ADOP data should be equal to DIOC, namely: $MIG^{DIOC}_{s,2000} = MIG^{ADOP}_{s,2000} * w_{s,2000}$
2. we re-weight the 1990 ADOP bilateral stocks using the weights for the 2000, namely $MIG^{ADOP_w}_{s,1990} = MIG^{ADOP}_{s,1990} * w_{s,2000}$
3. for bilateral stocks where we could not compute the weight due to missing values or zeroes, we compute the bilateral skill specific growth rate between 2000 and 2010 with DIOC and we assume a constant linear trend also for 1990. If we define $g_{s,00}$ the bilateral skill specific growth rate, then we can compute the missing bilateral skill migration stocks in the 1990 as follow: $MIG^{ADOP_w}_{s,1990} = MIG^{DIOC}_{s,2000}/(1 + g_{s,00})$.

After this procedure we have a panel of 25 destination countries with 194 origin countries. However, due to missing information in the 1990 for Luxembourg, Czech Republic and Slovak Republic on our ECI data, our final panel will be on 22 destination countries.

2.8.2 Additional Tables

Table A1: Correlates across different measures of Diversity

Diversity Measures	(1) Div_{HS}^{Mig}	(2) Mig_{HS}	(3) Div_{LS}^{Mig}	(4) Mig_{LS}
$N\#Origin$	0.263***	0.293***	0.264***	0.341***
Pol_{HS}^{Mig}	-0.386***	-0.0845	-0.343***	-0.173*
Pol_{LS}^{Mig}	0.0269	-0.181*	-0.160*	-0.144*
$Theil_{HS}^{Mig}$	-0.901***	-0.118	-0.775***	-0.177*
$Theil_{LS}^{Mig}$	-0.761***	-0.217**	-0.911***	-0.180*
$Theil_{HS}^{W,Mig}$	-0.639***	0.125	-0.567***	0.116
$Theil_{LS}^{W,Mig}$	-0.465***	0.0119	-0.646***	0.0979
$Theil_{HS}^{B,Mig}$	-0.234***	-0.235***	-0.184**	-0.281***
$Theil_{LS}^{B,Mig}$	-0.254***	-0.229**	-0.208**	-0.285***

Note: Correlations * p<0.1, ** p<0.05, *** p<0.01.

Table A2: OLS regression on Economic Complexity Index - Different Diversity Indexes

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable: *ECI*	OLS	OLS	OLS	OLS	OLS	OLS
Sample:	2^{nd} *Tercile*					
Diversity Measures	Div^{Mig}	Div^{Mig}	Pol^{Mig}	$Theil^{Mig}$	$Theil^{W,Mig}$	$Theil^{B,Mig}$
HS Index	0.128**	0.127**	-0.042	-0.142*	-0.107	-0.283**
	(0.058)	(0.059)	(0.054)	(0.072)	(0.074)	(0.123)
LS Index	0.024	0.025	0.092	0.055	0.085	0.256*
	(0.087)	(0.088)	(0.054)	(0.077)	(0.079)	(0.135)
Mig_{HS}	-0.072	-0.072	-0.048	-0.064	-0.073	-0.056
	(0.085)	(0.085)	(0.071)	(0.094)	(0.084)	(0.097)
Mig_{LS}	0.090	0.088	0.105	0.086	0.112	0.078
	(0.098)	(0.099)	(0.094)	(0.110)	(0.111)	(0.115)
N#Origin		0.000				
		(0.001)				
Observations	67	67	67	67	67	67
Countries	44	44	44	44	44	44
Adj. R-Square	0.34	0.37	0.21	0.24	0.15	0.19
Controls	✓	✓	✓	✓	✓	✓
Regional FE	✓	✓	✓	✓	✓	✓

Note: authors' calculations using data from Hausmann *et al.* (2014) and Alesina et al. (2016). Standard errors are clustered at country level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is a standardized measure of the Economic Complexity Index. Each regression includes the full set of controls of Table 3 of Alesina et al. (2016) plus the logarithm of GDP. The sample of of countries is composed by countries belonging to the 2^{nd} tercile of the ECI distribution. Each column shows different approaches to measure skill-specific birthplace diversity among immigrants: fractionalization index (col. 1), fractionalization index plus the number of migrants' countries of origin (col. 2), polarization index (col. 3), theil index (col. 4) and its decomposition: the within component (col. 5) and between component (col. 6).

Table A3: PPML regression on bilateral migration stocks like AHR

	(1) PPML	(2) PPML	(3) PPML
Dep var.	MIG_{All}	MIG_{LS}	MIG_{HS}
Pop destination 1960	0.389***	0.374***	0.443***
	(0.133)	(0.126)	(0.153)
Bilateral weighted distance	-3.367***	-3.484***	-3.414***
	(0.838)	(0.818)	(1.105)
Colonial relationship	1.607***	1.578***	1.908***
	(0.292)	(0.321)	(0.301)
Common ethnic language	1.854***	1.688***	2.388**
	(0.678)	(0.571)	(0.958)
Common official language	-0.193	-0.177	-0.319
	(0.585)	(0.555)	(0.729)
Common border	1.827***	1.934***	1.008***
	(0.237)	(0.245)	(0.173)
Horizontal Time difference	0.300***	0.229***	0.454***
	(0.113)	(0.084)	(0.164)
Observations	71060	71060	71060
Countries	190	190	190
R-Square	0.42	0.49	0.19
Year FE	✓	✓	✓
Origin*Year FE	✓	✓	✓

Note: authors' calculations on Artuc et al. (2015) and Alesina et al. (2016) data.

Table A4: Descriptive Statistics Migration

	(1)	(2)	(3)	(4)	(5)
Variable	Obs	Mean	Std. Dev.	Min	Max
Div_{All}^{Mig}	200	0	1	-3.396	1.026
Div_{HS}^{Mig}	200	0	1	-3.955	0.827
Div_{LS}^{Mig}	200	0	1	-3.266	1.077
$Div_{All}^{Mig,Grav}$	200	0	1	-4.034	0.943
$Div_{HS}^{Mig,Grav}$	200	0	1	-5.968	0.664
$Div_{LS}^{Mig,Grav}$	200	0	1	-3.744	0.976
$Div_{All}^{Mig,SS}$	200	0	1	-3.225	1.118
$Div_{HS}^{Mig,SS}$	200	0	1	-3.283	1.044
$Div_{LS}^{Mig,SS}$	200	0	1	-3.156	1.163
Mig_{All}	200	0	1	-0.594	4.712
Mig_{HS}	200	0	1	-0.625	5.551
Mig_{LS}	200	0	1	-0.587	4.910
Mig_{All}^{Grav}	200	0	1	-0.594	4.850
Mig_{HS}^{Grav}	200	0	1	-0.638	5.440
Mig_{LS}^{Grav}	200	0	1	-0.578	4.831
Mig_{All}^{SS}	200	0	1	-0.689	6.0108
Mig_{HS}^{SS}	200	0	1	-0.293	10.299
Mig_{LS}^{SS}	200	0	1	-0.633	6.343

Note: authors' calculations on Özden et al. (2011), Alesina et al. (2016) and ADOP data.

Table A5: Correlations across birthplace diversity indices

	(1) Div_{All}^{Mig}	(2) Div_{HS}^{Mig}	(3) Div_{LS}^{Mig}	(4) $Div_{All}^{Mig,Grav}$	(5) $Div_{HS}^{Mig,Grav}$	(6) $Div_{LS}^{Mig,Grav}$	(7) $Div_{All}^{Mig,SS}$	(8) $Div_{HS}^{Mig,SS}$
Div_{HS}^{Mig}	0.869***							
Div_{LS}^{Mig}	0.993***	0.841***						
$Div_{All}^{Mig,Grav}$	0.337***	0.347***	0.348***					
$Div_{HS}^{Mig,Grav}$	0.108	0.174*	0.108	0.708***				
$Div_{LS}^{Mig,Grav}$	0.358***	0.350***	0.372***	0.991***	0.639***			
$Div_{All}^{Mig,SS}$	0.685***	0.584***	0.700***	0.224**	0.0622	0.239***		
$Div_{HS}^{Mig,SS}$	0.637***	0.637***	0.645***	0.244***	0.0722	0.258***	0.927***	
$Div_{LS}^{Mig,SS}$	0.683***	0.559***	0.700***	0.209**	0.0495	0.226**	0.994***	0.890***

Note: authors' calculations on Alesina et al. (2016), Artuc et al. (2015) and Özden et al. (2011) data. This table show correlations among variables with their level of significance * p<0.1, ** p<0.05, *** p<0.01.

Table A6: Correlations across shares of migrants

	(1) Mig_{All}	(2) Mig_{HS}	(3) Mig_{LS}	(4) Mig_{All}^{Grav}	(5) Mig_{HS}^{Grav}	(6) Mig_{LS}^{Grav}	(7) Mig_{All}^{SS}	(8) Mig_{HS}^{SS}
Mig_{HS}	0.869***							
Mig_{LS}	0.996***	0.836***						
Mig_{All}^{Grav}	0.980***	0.871***	0.972***					
Mig_{HS}^{Grav}	0.794***	0.834***	0.771***	0.824***				
Mig_{LS}^{Grav}	0.963***	0.851***	0.959***	0.987***	0.774***			
Mig_{All}^{SS}	0.811***	0.663***	0.816***	0.798***	0.722***	0.796***		
Mig_{HS}^{SS}	0.0208	0.0362	0.0218	0.0112	0.235***	0.00274	0.223**	
Mig_{LS}^{SS}	0.810***	0.651***	0.818***	0.808***	0.665***	0.840***	0.963***	0.137

Note: authors' calculations on Alesina et al. (2016), Artuc et al. (2015) and Özden et al. (2011) data. This table show correlations among variables with their level of significance * p<0.1, ** p<0.05, *** p<0.01.

Table A7: Alternative IV Estimates - Shift-share based on 1960 distribution

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Variable: ECI									
Panel A - Overall Sample									
Div_{All}^{Mig}	0.296***	0.116**	0.294***						
	(0.091)	(0.059)	(0.090)						
Mig_{All}	0.011	-0.014	0.006						
	(0.067)	(0.065)	(0.066)						
Div_{HS}^{Mig}				0.361***	0.110**	0.355***			
				(0.119)	(0.054)	(0.122)			
Mig_{HS}				0.052	-0.064	-0.022			
				(0.091)	(0.135)	(0.141)			
Div_{LS}^{Mig}							0.279***	0.109*	0.286***
							(0.086)	(0.057)	(0.085)
Mig_{LS}							0.018	0.020	0.038
							(0.066)	(0.073)	(0.072)
Observations	200	200	200	200	200	200	200	200	200
Countries	100	100	100	100	100	100	100	100	100
K-P rk Wald F-stat	22.75	425.25	11.57	17.08	29.41	9.08	27.65	126.61	14.05
Adj. R-Square	0.73	0.75	0.73	0.71	0.75	0.71	0.73	0.75	0.73
Panel B - 2^{nd} Tercile									
Div_{All}^{Mig}	0.108	0.176***	0.117*						
	(0.067)	(0.048)	(0.068)						
Mig_{All}	0.067	0.096*	0.085						
	(0.049)	(0.051)	(0.052)						
Div_{HS}^{Mig}				0.074	0.147***	0.038			
				(0.070)	(0.036)	(0.082)			
Mig_{HS}				0.009	-0.066	-0.060			
				(0.045)	(0.058)	(0.068)			
Div_{LS}^{Mig}							0.106	0.159***	0.123*
							(0.066)	(0.052)	(0.070)
Mig_{LS}							0.072	0.112**	0.102*
							(0.050)	(0.056)	(0.058)
Observations	67	67	67	67	67	67	67	67	67
Countries	44	44	44	44	44	44	44	44	44
K-P rk Wald F-stat	20.88	517.43	9.53	8.44	15.95	2.85	20.30	626.49	8.22
Adj. R-Square	0.28	0.30	0.28	0.30	0.33	0.20	0.26	0.27	0.26
Instr *Div*	✓		✓	✓		✓	✓		✓
Instr *Mig*		✓	✓		✓	✓		✓	✓

Note: authors' calculations using data from Hausmann *et al.* (2014), Alesina et al. (2016) Özden et al. (2011), and Artuc et al. (2015) data. Standard errors are clustered at country level. * p<0.1, ** p<0.05, *** p<0.01. This table shows the effect of birthplace diversity on our standardized measure of Economic Complexity,after controlling for the full set of controls of Table 3 of Alesina et al. (2016) and for the logarithm of real GDP per capita. Birthplace diversity is instrumented with predicted stocks through shift-share methodology based on the 1960 distribution, while the migration share is instrumented with predicted stocks from gravity model.

Table A8: Correlations of diversity and alternative diversification measures

	(1) ECI	(2) $ln(GDPpc)$	(3) Div_{All}^{Mig}	(4) Div_{HS}^{Mig}	(5) Div_{LS}^{Mig}	(6) Gini on Export	(7) MaxMin	(8) Log Var	(9) Theil Index	(10) Theil Index (w)
ECI	1									
$ln(GDPpc)$	0.720***	1								
Div_{All}^{Mig}	0.338***	0.447***	1							
Div_{HS}^{Mig}	0.209**	0.295***	0.869***	1						
Div_{LS}^{Mig}	0.326***	0.427***	0.993***	0.841***	1					
Gini on Export	-0.844***	-0.636***	-0.343***	-0.218**	-0.331***	1				
MaxMin	-0.588***	-0.352***	-0.129	-0.0480	-0.107	0.629***	1			
Log Var	-0.724***	-0.485***	-0.216**	-0.101	-0.197**	0.836***	0.923***	1		
Theil Index	-0.785***	-0.545***	-0.248***	-0.119	-0.227**	0.880***	0.896***	0.976***	1	
Theil Index (w)	-0.697***	-0.360***	-0.156*	-0.105	-0.141*	0.812***	0.873***	0.938***	0.925***	1
Theil Index (b)	-0.623***	-0.660***	-0.315***	-0.0987	-0.293***	0.641***	0.569***	0.643***	0.723***	0.407***

Note: authors' calculations using data from Hausmann *et al.* (2014) and Artuc et al. (2015). This table show correlations among variables with their level of significance * p<0.1, ** p<0.05, *** p<0.01.

Table A9: OLS regression on Countries Top Products Ubiquity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Variable: *PCI*	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Products	*Top 3*			*Top 5*			*Top 10*		
Div_{All}^{Mig}	0.034			0.040			0.071*		
	(0.055)			(0.046)			(0.042)		
Mig_{All}	-0.073			-0.014			0.027		
	(0.055)			(0.055)			(0.043)		
Div_{HS}^{Mig}		0.035			0.027			0.063	
		(0.055)			(0.047)			(0.039)	
Mig_{HS}		-0.084			0.004			0.041	
		(0.070)			(0.062)			(0.049)	
Div_{LS}^{Mig}			0.027			0.035			0.068
			(0.052)			(0.044)			(0.042)
Mig_{LS}			-0.064			-0.011			0.027
			(0.055)			(0.054)			(0.042)
$ln(GDPpc)$	0.118	0.118	0.115	0.152	0.152	0.153	0.151	0.160*	0.153
	(0.133)	(0.129)	(0.132)	(0.126)	(0.121)	(0.125)	(0.101)	(0.095)	(0.101)
Observations	200	200	200	200	200	200	200	200	200
Countries	100	100	100	100	100	100	100	100	100
Adj. R-Square	0.49	0.49	0.49	0.58	0.58	0.58	0.69	0.69	0.69
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Regional FE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations using data from Hausmann *et al.* (2014) and Alesina et al. (2016). Standard errors are clustered at country level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is the country average Product Complexity index over the top 3, top 5 and top 10 products in terms of competitive export. Each regression includes the full set of controls of Table 3 of Alesina et al. (2016).

Table A10: OLS regression on Diversification and Ubiquity Measures

	Maxmin			*Log Var*			*Gini Index on Export*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Div_{All}^{Mig}	-0.014			-0.129			-0.010***		
	(0.020)			(0.080)			(0.003)		
Mig_{All}	-0.021			-0.104			0.000		
	(0.020)			(0.077)			(0.003)		
Div_{HS}^{Mig}		-0.014			-0.121			-0.012***	
		(0.019)			(0.075)			(0.003)	
Mig_{HS}		-0.029			-0.125			0.000	
		(0.032)			(0.121)			(0.004)	
Div_{LS}^{Mig}			-0.011			-0.111			-0.010***
			(0.019)			(0.078)			(0.003)
Mig_{LS}			-0.020			-0.099			-0.000
			(0.019)			(0.073)			(0.003)
$ln(GDPpc)$	0.039	0.038	0.036	0.071	0.049	0.059	-0.008*	-0.009*	-0.009*
	(0.030)	(0.032)	(0.029)	(0.120)	(0.128)	(0.119)	(0.005)	(0.005)	(0.005)
Observations	200	200	200	200	200	200	198	198	198
Countries	100	100	100	100	100	100	100	100	100
Adj. R-Square	0.48	0.48	0.48	0.62	0.62	0.61	0.83	0.83	0.83

	Theil Index			*Theil Index Within*			*Theil Index Between*		
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Div_{All}^{Mig}	-0.130			-0.103			-0.026		
	(0.089)			(0.077)			(0.030)		
Mig_{All}	-0.079			-0.071			-0.002		
	(0.084)			(0.078)			(0.039)		
Div_{HS}^{Mig}		-0.106			-0.149**			0.045	
		(0.085)			(0.068)			(0.029)	
Mig_{HS}		-0.110			-0.107			-0.001	
		(0.136)			(0.115)			(0.040)	
Div_{LS}^{Mig}			-0.115			-0.088			-0.026
			(0.086)			(0.075)			(0.028)
Mig_{LS}			-0.075			-0.067			-0.001
			(0.081)			(0.076)			(0.038)
$ln(GDPpc)$	0.048	0.028	0.038	0.079	0.097	0.068	-0.037	-0.071	-0.038
	(0.133)	(0.143)	(0.132)	(0.113)	(0.115)	(0.112)	(0.059)	(0.053)	(0.058)
Observations	198	198	198	198	198	198	200	200	200
Countries	100	100	100	100	100	100	100	100	100
Adj. R-Square	0.70	0.70	0.70	0.57	0.58	0.57	0.70	0.70	0.70
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Regional FE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations using data from Hausmann *et al.* (2014) and Alesina et al. (2016). Standard errors are clustered at country level. * p<0.1, ** p<0.05, *** p<0.01. This table shows the effect of birthplace diversity on different measures of product diversification: Max min on export (col. (1)-(3)), Log Var on Export (col. (4)-(6)), Gini index on Export (col. (7)-(9)), Theil Index (col. (10)-(12)), Theil Index within (col. (13)-(15)) and Theil Index between (col. (16)-(18)).

Chapter 3

Skill of the Immigrants and Vote of the Natives: Immigration and Nationalism in European Elections 2007-2016

Abstract[1]

In this paper we analyze the impact of local immigration on political preferences for "nationalistic" ideas, as expressed by citizens' votes in 28 national elections in twelve EU-15 countries, between 2007 and 2016. We derive the degree of "nationalism" of each party's political agenda, using data from the Manifesto Project Database, and construct an index of individual nationalist voting in a region across elections and countries. We then analyze how the change in this index was associated with the change in local immigration, using an instrumental variable strategy. The instrument interacts immigrant settlements as of 2005 with the skill composition of recent immigrant flows. IV estimates show that larger inflows of *highly educated* immigrants in a region were associated with a *decrease* in the nationalism index. To the contrary, the inflow of *less educated* immigrants was associated with a vote shift that increased the nationalism index. These effects (in both directions) were stronger for less educated citizens. The aggregate results derive in part from an individual shift towards nationalism associated with less-skilled immigration, greater participation of young voters, and pro-Europe attitudes associated with high-skilled immigration. Simulations based on the estimated coefficients show that on average immigration policies balancing the number of

[1]This paper is coauthored with Giovanni Peri and Simone Moriconi

high-skilled and less-skilled immigrants would be associated with a decline in the nationalist index in almost all European regions.

3.1 Introduction

Recent elections in several European countries (UK in 2016, France, Germany and Austria in 2017, Italy in 2018) and in the US (in 2016) have shown a common trend: nationalist parties and candidates with a strong anti-immigrant rhetoric have gained votes and have often been elected to office. The common narrative is that these parties have expressed the fear, resentment and sometimes anger of their constituents in response to the increased number of immigrants in their communities. The perception of immigrants as outsiders who compete for jobs, drain public finances and constitute a threat to national values, especially if they are from different cultures, has fueled this sentiment.

Politicians from populist parties have played into these fears. They tend to overemphasize immigrants' economic and fiscal cost, and their threat to public order and propensity to criminal activity. They usually portray immigrants as impoverished, unskilled, unproductive, and hence a drain on the economy. Examining data from recent decades of immigration, however, tells a different story. The data and related academic research show that a very large percentage of immigrants to Europe and the US in recent decades were highly-educated and skilled. Often, the immigrant population contains a larger share who are college-educated than the native population (see Docquier *et al.*, 2014). Related to this, economists find that immigrants, especially highly educated ones, contribute to the performance and growth of the receiving economy (Borjas, 2019). Immigrants have enhanced technological and scientific innovation (Kerr and Lincoln, 2010), boosted productivity (Peri, 2012) and produced, on average, a fiscal surplus (Orrenius, 2017). In general, immigrants are a combination of highly- and less-educated individuals, and have an overall positive effect on a country's aggregate economic growth and fiscal balance. The economic and fiscal effect of immigrants in specific cases, however, depends on their characteristics and on the features of the receiving economy.

European voters likely mis-perceive the average contribution of immigrants (as shown in Alesina *et al.*, 2018) and often undervalue their positive contributions and overvalue their costs. But does the increased exposure to immigrants elicit an electoral response? And does the inflow of different types of immigrants produce different responses? What characteristics affect citizens voting-response to immigration? In this paper, we address these questions. We use a sample of elections in 12 EU-15 countries, over the period 2007-2016 for which

we have a complete coverage of elections and detailed data on immigrants. To identify the effect on votes we exploit the large differential variation of immigrants' population share at the regional level within those countries, driven mainly by changes in education levels and countries-of-origin for recent flows of immigrants.

The overarching question asked in this paper, and in a series of recent papers which we will review below, is about how immigration has affected the political vote of natives, either by changing their attitudes or their willingness to vote. A previous branch of research in economics and political science has focused on a related question, namely: what determines individual attitudes and policy preferences toward immigrants and immigration? And do these differ towards high- and less-skilled immigrants? Several of those papers looked at whether theories of labor market competition (Mayda, 2006; Scheve and Slaughter, 2001; Hainmueller and Hiscox, 2007) or fiscal effects of immigrants (Facchini and Mayda, 2009; Hanson *et al.*, 2007) could explain individual attitudes towards immigrants of different types. The labor market competition theory predicts stronger opposition from natives with skills similar to that of the majority of immigrants. The fiscal theory predicts stronger opposition from richer people who will pay the redistributive costs. These studies find at best mixed support for both explanations.

One robust finding of these studies is that a voters' schooling level is a very strong predictor of her attitude towards immigration. College education is associated with positive attitudes towards immigrants (of all types) and approval for more open immigration policies. Hainmueller and Hiscox (2010) tackles the question of what determines natives' attitudes towards high/less-skilled immigrants and how this depends on their education levels. Their findings, based on data from a survey of 2,285 people, show that neither labor market competition nor fiscal competition explain those attitudes well. In fact, more-educated natives have more-positive attitudes toward both types of immigrants, relative to less-educated natives. This is at odds with the idea that individuals are opposed to immigrants with similar labor market skills and welcome immigrants with different skills. Moreover they find that richer people are more in favor of immigrants than poorer people, and this is even stronger if they live in states with more redistribution. Hainmueller and Hiscox (2010) argue that economic self-interest does not motivate the vote; rather, a perception of what is good for society does. This would explain the more-favorable attitudes toward high-skilled immigrants who, arguably, have a larger positive effect on the overall economy and society. An alternative "cultural" explanation, in which tolerance and preference for cultural diversity is associated with higher education levels, is also consistent with those findings. Card *et al.* (2012) show that the individual "cultural" assessment of immigrants drives overall opinion about immigration, while

economic evaluations of immigrants' contributions are less-relevant. Poutvaara and Steinhardt (2018) show that the individual perception of not having what one deserves in life is strongly associated with worries about immigration, particularly among the less-skilled.

A more recent strand of the literature has directly tested the link between immigration and the electoral vote. The specific question asked in several papers is whether an increased inflow of immigrants affected the vote of natives and in what direction. The common observation of higher immigration flows and growing electoral consensus for anti-immigration parties has driven scholars to inquire after the extent to which the first determines the second. This effect requires that changes in the local presence of immigrants either changes attitudes for (at least) some part of the native population, or changes their willingness to vote. Whatever is their initial attitude towards immigrants, natives either become more concerned with this issue, and more aware that this is relevant to them (i.e. the issue gains salience), or they change their opinions of immigrants in response to the inflow. In-line with the literature on "political persuasion" (e.g. Della Vigna and Gentzkow, 2010), some groups may revise their views of immigrants in a Bayesian updating (contact theory, prevalence of personal experience). In this process, individuals acquire "new information" from meeting or interacting with immigrants, or from being exposed to news related to immigrants in their local community. This may change the vote of some groups of individuals, and the platforms of elected parties–in terms of immigration policies–would reflect such a shift. Some people may not use the new pieces of information to correct but rather to confirm their biased political priors (see e.g. Nyhan and Reifler, 2010). In this perspective, the key question is not what personal characteristics affect people's views of immigrants, but rather what characteristics of people are associated with larger or smaller revisions of their preferences for immigrants as a consequence of immigration. Similarly, if this is correct, it is of interest which characteristics of immigrants produce a positive or negative revision of such preferences.

A series of recent studies has taken this approach, focusing on how immigration has changed the political landscape in individual countries, especially in terms of vote shares received by populist/nationalistic parties.[2] These studies have focused on individual countries and the relationship between immigrant shares of the population and the vote for specific parties at the local level.[3] Each of these papers is focused on a specific country and the identification

[2]A few papers have looked at other determinants of the increase in the share of populist vote in many European countries. Algan *et al.* (2017), for instance analyzes how the intensity of the Great Recession affected those votes.

[3]Among them, Otto and Steinhardt (2014) analyze the effect of immigration on the vote for the German People's Union in Hamburg. Barone *et al.* (2016) look at immigration and vote to the center-right/Northern League coalition in Italy. Halla *et al.* (2017) look at votes to the Freedom Party of Austria. Harmon (2018) look at cultural distance of immigrants and vote for the Danish People's Party in Denmark. Brunner

is based on instrumental variable estimation, using historical location of immigrants and a shift-share instrumental-variable approach. The specific national context is a limitation of these papers as the generalization of the findings can be problematic. However, taken together, these studies suggest that an increase in the immigrant population share in a region/municipality is associated with a shift of the vote towards right-wing nationalist parties. The papers do not focus in particular on different effects of different group of immigrants.

The present paper generalizes the analysis of immigration and voting behavior by homogenizing data for all parties and all elections of 12 European countries over the 2007-2016 period. In order to compare votes to different parties in different countries we introduce an innovation: we classify each party on a common scale in terms of the intensity of their "nationalism". We quantify nationalism using textual analysis of their political agenda as described in their Political Manifesto and measured by the *Manifesto Project Database*. We merge the measure of nationalism intensity of a party (on a standardized scale) with information about votes of individuals from the *European Social Survey*, a large and representative sample of EU citizens in the 12 considered countries. We use this information to construct a panel of votes in 114 regions over 28 European elections, where the dependent variable is the "nationalism index" of such vote. We finally merge it with the change in high- and less-skilled immigrants as share of adult population for each region (calculated using the *European Labor Force Survey*, a survey representative of the EU population). Our main specification analyzes how the change in local immigrant share was associated with the "nationalism index" of voters in EU regions.

In order to go beyond correlations, we use an instrumental variable (IV) strategy and we construct a shift-share instrument (Card, 2001) to predict, separately, high-skilled and less-skilled immigration. The IV is based on the location of immigrants by area-of-birth in 2005 interacted with the growth of each immigrant group by area-of-origin and skill level. The differentiation in country-of-origin and skill-intensities of groups allows us to estimate two separate effects for highly-skilled and less-skilled immigrants. While the variation in the inflow of high- and less-skilled immigrants is not random across regions, the use of our instrument and the inclusion of a rich set of fixed effects and controls should reduce concerns about omitted variable bias. We base our instrument on the immigrant distribution in 2005.

and Kuhn (2018) analyze the vote for the Swiss People's Party in Switzerland. Edo *et al.* (2019) look at the French vote for the "Front National" in France. Mayda *et al.* (2018) show that, in the US, high-skilled immigration has been strongly associated with a higher share of Democratic vote and less-skilled immigration with a higher share of Republican vote. Finally, Tabellini (2018) shows that, despite bringing local economic prosperity, European immigration to US cities during 1910-1930 increased the vote for anti-immigration conservative legislators.

Moreover we test that the instrument is not correlated with previous changes in demographic and economic variables at the regional level in the period 2000-07. This test checks that correlations of the shift-share IV with pre-existing economic and demographic trends is not a source of concern, as suggested in some recent papers such as Jaeger *et al.* (2018).

An additional contribution of our paper is that, by using individual data on voters (and not aggregate district-level voting totals, as do most of the existing papers) we can look at heterogeneity in vote response. We can also test the impact on individual attitudes towards policies, politics and towards immigrants, to produce a complete picture of the effect of immigration on individual policy preferences. This paper, therefore, connects the studies that analyze immigration and individual attitudes (such as Hanson *et al.*, 2007 or Hatton, 2016, on refugees and asylum policies) with those analyzing immigration and voting behavior (e.g. Halla *et al.*, 2017; Barone *et al.*, 2016). By looking at the impact of immigrants on vote turnout, we can discuss whether a change in individual attitudes or a change in voting participation is the more likely channel for the aggregate effect of immigration on elections.

We show three main findings. First, we find a strong negative correlation between highly-educated immigrants (as a share of the population) and intensity of nationalist preferences at the regional level. We also find an opposite, significant positive association of the population share of less-skilled immigrants and nationalist preferences. Second, we find that the change in preferences, in response to high- and less-skilled immigrant share, are larger for less-educated citizen voters than for highly-educated citizen voters. The effect is also stronger for younger voters than for older voters. This is consistent with the idea that people with lower previous knowledge/exposure to immigrants respond more significantly to their presence, both in positive and negative terms. Third, we provide some insight into the channels through which immigration affects nationalist voting. One significant effect is the increased participation of young voters in response to increased high-skilled immigrant share. As high-skilled immigrants are more pro-Europe and less-populist, a peer effect may induce young citizen voters to participate more in the vote. This demographic group is the less-nationalist group of voters, with the most positive attitudes towards the EU, hence this is a channel of reduced support for nationalism. There is also some evidence of changes in individual opinion, in terms of the attitudes towards politics and towards immigrants. In particular, we show that an increase in high-skilled immigrants raises individuals' trust in legislative power both at the EU and the national level, while an increase in less-skilled immigrants raises natives' skepticism against immigrants, instead. This seems to indicate that interactions of natives with highly-educated immigrants bring individuals to revise their preferences in favor of them. However, interactions with less-educated immigrants produce the opposite

effect. This updating of preferences implies that changes in individual preferences affect political opinions. A recent paper by Alesina *et al.* (2018) analyzes how individual views about immigration affect the preferences of citizens for redistribution. The paper shows, interestingly, that less-educated people in sectors with high levels of immigrant competition tend to overestimate the presence and the cost of immigrants. This could be a reason for the fact that we find the pro-nationalist effect of less-skilled immigration is larger on less-educated natives than on the more educated ones.

We use our findings to simulate possible political scenarios of nationalism in Europe. Simulations suggest that being most sensitive to the skill-mix of both migrants and natives, nationalist voting is most likely to be attenuated by an increase in the relative amount of high-skilled migration compared to less-skilled migration, and by an increase in the share of tertiary-educated natives. They also show that the composition of migration flows is key to reduce nationalism: restrictions of less-skilled migration are not strictly necessary, as nationalist voting would equally shrink if inflows of high-skilled migrants fully balance less-skilled migration.

The rest of the paper is organized as follows. In Section 3.2 we discuss the data and, specifically, we document the intensity of nationalistic preferences in recent elections and the increase in migrants as shares of local resident populations. In Section 3.3 we present the empirical strategy, first, and then our instrumental variable approach and validation. In Section 3.4 we discuss the baseline results on immigration and nationalist voting. In Section 3.5 we present our results by exploiting individual- and regional level heterogeneity. In Section 3.6 we decompose the channels of nationalist voting behavior, focusing on individual opinion and turnout. In Section 3.7, we present the simulations. Section 3.8 concludes the paper.

3.2 Data and Summary Statistics

In order to analyze the change in "nationalism" of European citizens in response to the inflow of immigrants, we combine three different datasets: The European Social Survey (ESS), the Manifesto Project Database (MPD) and the European Labor Force Survey (EULFS). Our primary data source is the European Social Survey (ESS). This was administered in 8 waves (one every two years) in 36 countries between 2002 and 2016. The ESS selects a random sample of individuals which is representative of the national population over 18 in each country. On average, each wave contains around 1,500 individuals for each country. The

data include detailed information on personal and family characteristics such as age, gender, education, marital status, number of children in the family, place of birth and labor market characteristics such as employment status, hours worked and occupation. It also includes detailed information on parental background, such as parents' education, employment status, occupation when the respondent was 14 years old, and their own country of birth.

ESS data report also two specific questions on voting: (i) "*which party did you vote in the last national election?*", and (ii) "*which party do you feel close to?*". The individuals respond by identifying party names, and we link them to information on their political agenda obtained from the second database, the *Manifesto Project Database (MPD)* (see Budge *et al.*, 2001; Klingemann et al., 2006). This MPD describes the political manifesto of 1,093 parties over 715 parliamentary elections covering all the countries and the years we consider.[4] Each party's political manifesto is analyzed through content analysis. Specifically, the share of quasi-sentences related to each political topic as a fraction of all sentences in the manifesto is taken as a measure of the relevance of that topic in the party's political agenda. Moreover, for almost each topic, the share of favorable/positive and unfavorable/negative mentions are available. This implies that we can construct two measures: a measure of *preference* for the topic, computed as the difference between positive and negative mention shares, and one of *salience* of the topic, equal to the sum of shares of positive and negative mentions. Both are indices, with the first ranging between 0 and 1 and the second between -1 and +1. We will use the "preference" measure in our analysis. Results obtained using the "salience indicator" are usually similar and are shown in the Appendix (results available in Table A2). The indicators described above are typically used in the Political Sciences Literature (see, for instance, Budge *et al.*, 2001; Klingemann et al., 2006).

The MPD covers several topics, ranging from protectionism, to the role of traditional morality, civic mindedness, multiculturalism and more. In our analysis we focus on parties' preferences on two topics: (i) The European Community/European Union and (ii) National way of life. The first topic includes mentions of the EU, the desirability of its expansion, increasing its influence, and competences. The second include all mentions of nationalism, patriotism, pride of citizenship, and focus on the nation. For each party we construct such indices for the first year the party enters the dataset, for the last year in the dataset, and for the average over all years of presence in the data. These indicators will capture the initial, average and final position of the party on the issue,[5] We consider the"preference" indicator for each of

[4]The MPD includes all parties that participated to national elections and obtained at least one seat in their parliament over the 1945-2017 period, covering all democratic countries in the OECD and Eastern Europe.

[5]The MPD tracks changes in parties' names, as well as parties' fusions and splits. The list of the parties

the two issues (EU and national way of life) and we combined them using a Principal Component Analysis (PCA), summarizing them in what we call the *Intensity* of "nationalism". This correlates negatively with favorable mentions of the EU variable, and positively with favorable mentions of the national way of life. Results of the PCA are available in the on-line 3.9.3. Finally, we link this index to the individual voter using the reported party voted-for in the last election from the ESS.

ESS data includes information on the location of respondents at the regional (NUTS2) level for all EU countries with a few exceptions (e.g. Austria, Germany, UK) where the local units identified are larger (NUTS1 level).[6] The geographic location of the ESS respondent allows us to link their data with regional variables in a given year. In particular, we calculate regional variables aggregating individual data from our third dataset, the *European Labor Force Survey (EULFS)* which reports information on age, employment status and education, for the population 15 years and older and is representative of the European population at the regional level. Starting from 2005, EULFS includes foreign-born individuals and reports for each of them their birthplace, distinguishing fifteen different regions of origin.[7] We use this variable to compute the share of immigrants in the total population in 2005 at the regional level as follows:

$$m^{O}_{r,s,t} = \frac{M^{O}_{r,s,t}}{Pop_{r,2005}}, \tag{3.1}$$

where $m^{O}_{r,s,t}$ is the share of migrants and $M^{O}_{r,s,t}$ is the total stock of migrants in region r born in the group of origin countries O among $\{All, Non\ EU28,\ EU28\}$, with skills $s \in \{All, HS, LS\}$ at time t.

We combine the individual ESS samples with EULFS migration data aggregated at the regional level and MPD data on the political agenda of parties. Our data includes only countries for which at least two elections are included in the 2007-2016 period, and that received significant immigrant flows. This leaves eleven EU-15 countries,[8] plus Switzerland as our sample for the 2007-2016 period. The countries in our final sample are Austria, Belgium,

included in the analysis is available in Table A13 in the on-line 3.9.4.

[6]The "Nomenclature for Territorial Units for Statistics","NUTS" system, partitions EU countries into Macro-Regions, Regions and Provinces which are called NUTS1, NUTS2 and NUTS3 level respectively. As stated in the text we use the intermediate, NUTS2, level commonly indicated as "region" in our analysis.

[7]These are the residence country, a foreign EU15 country, a foreign EU member state entered in the EU on 2004, a foreign EU Accession country in 2007/2013, EFTA, Other Europe, North Africa, Other Africa, Near and Middle East, East Asia, South and South East Asia, North America, Central America and Caribbean, South America and Australia and Oceania.

[8]We had to exclude Italy and Greece from the analysis, as the ESS data during 2007-2016 period covered only one election in each of these countries (2013 in Italy, and 2009 in Greece).

Denmark, Finland, France, Germany, Ireland, Portugal, Spain, Switzerland, Sweden and the United Kingdom. For these countries, we construct the origin-skill-specific migrants shares of the total population in 2005, at the regional level.[9] The sample covers 114 regions; for these regions we also obtain regional data from Eurostat and the OECD, covering GDP per capita, unemployment rate, percentage of tertiary-educated individuals, value of social benefits other than social transfers in-kind, the ratio of 0-14 to 15-65 years old population, and a measure of total number of crimes.[10]

Table 3.1: Elections and ESS Rounds by Country and Year

	(1)	(2)	(3)	(4)
Country	# Elections	Election Years	# Survey Rounds	Survey Years
Austria	2	2008, 2013	3	2010, 2014, 2016
Belgium	2	2010, 2014	4	2010, 2012, 2014, 2016
Denmark	2	2007, 2011	3	2010, 2012, 2014
Finland	3	2007, 2011, 2015	4	2010, 2012, 2014, 2016
France	2	2007, 2012	4	2010, 2012, 2014, 2016
Germany	2	2009, 2013	4	2010, 2012, 2014, 2016
Ireland	2	2011, 2016	4	2010, 2012, 2014, 2016
Portugal	3	2009, 2011, 2015	4	2010, 2012, 2014, 2016
Spain	3	2008, 2011, 2016	4	2010, 2012, 2014, 2016
Sweden	2	2010, 2014	4	2010, 2012, 2014, 2016
Switzerland	3	2007, 2011, 2015	4	2010, 2012, 2014, 2016
United Kingdom	2	2010, 2015	4	2010, 2012, 2014, 2016

Note: Column (1) shows the number of elections available from ESS and column (2) the year of each elections. Column (3) shows the number of ESS rounds after 2010 by country and column (4) the year of each round. Source: ESS.

The main source of identifying variation in our empirical analysis is the within-region changes in voting behavior between different national elections in a country. In fact, as ESS data convey information on the vote each respondent casted at the most recent national election, rounds ESS5-ESS8 (carried out between 2010 and 2016), report the votes of respondents at elections held during 2007-2016. Moreover, if more than one round was conducted between two consecutive elections, the respondents to these waves would have all voted to the same election i.e. their vote, recorded in different waves, actually refers to the same election. Thus, in our empirical analysis, we ultimately exploit the overtime variation between electoral

[9]Germany provides information on the total stock of the foreign born population and their nationality, but not on their birthplace. We use the nationality of immigrants to approximate their birthplace.

[10]An extended data section describing these variables in greater detail is available in the on-line 3.9.2 and Table A5.

episodes within a country. This is better understood by looking at Table 3.1, which reports the number and year of elections covered by ESS for each country in columns (1) and (2) and the number and years of the surveys in columns (3) and (4). For instance, in the case of France, we record/observe the *most recent* voting behavior by respondents to waves conducted in 2010, 2012, 2014 and 2016 (see column (4)). Accordingly, we allocate these votes to the national elections of 2007 and 2012, respectively, as these are the last ones in which these survey respondents may have participated (see column (2) in Table 3.1). To be more precise, we assign the individual votes recorded in the 2010 and 2012 ESS waves to the 2007 French election, as this was the most recent national election for these survey respondents. Similarly, we assign votes recorded in the 2014 and 2016 ESS waves to the 2012 election.[11] In this way, we map the variation between elections (column (2) in Table 3.1) into variation between survey years (column (4)). We exploit this source of variation for the baseline specification, where we pool the individual data in repeated representative cross sections. We include region fixed effects. As the explanatory variable of interest is the regional share of foreign-born, the variation that we exploit is a region-level pseudo panel.

We are able to extend the analysis further, and construct from individual data a genuine short panel with two observations over time for each individual. To do so, we combine the retrospective information on the votes at the last national elections (as used above) with the party preferences at the time of the survey, revealed by the question "What party do you feel close to?". For each individual, we consider as the first data point the party she/he voted for at the last election, as discussed above. We then consider as another data point the party preference expressed at the time of the survey which is available each survey year. Hence, we have two observations for each individual, over time, and we can use these data to analyze the individual evolution of party preference in a longitudinal approach, controlling for individual fixed effects. The limitation of this approach is that the question about the current party has many missing values and the non-response rate may be non-random in terms of party affiliation. Table A6 in the on-line Appendix provides descriptive statistics for the sample of individuals who reveal their current party preference. Table A7 provides a measure of differences in terms of political preferences between the party voted-for at the last national elections and the currently-preferred party. On average, 16% of the sampled individuals changed party preference between the last national election and the year of the survey. While the database using repeated cross-section of electoral votes will be used to perform the main analysis of the vote across regions over time, the longitudinal dataset

[11] When the survey and election years correspond (e.g. in France in 2012, or Sweden in 2010 and 2014), we use the exact dates of the interviews (i.e. including months and days if needed), which are available in the ESS, to determine which is the most recent national elections in which each interviewee had participated.

allows a window into the change in individual political preferences over time.

Table 3.2: Summary Statistics

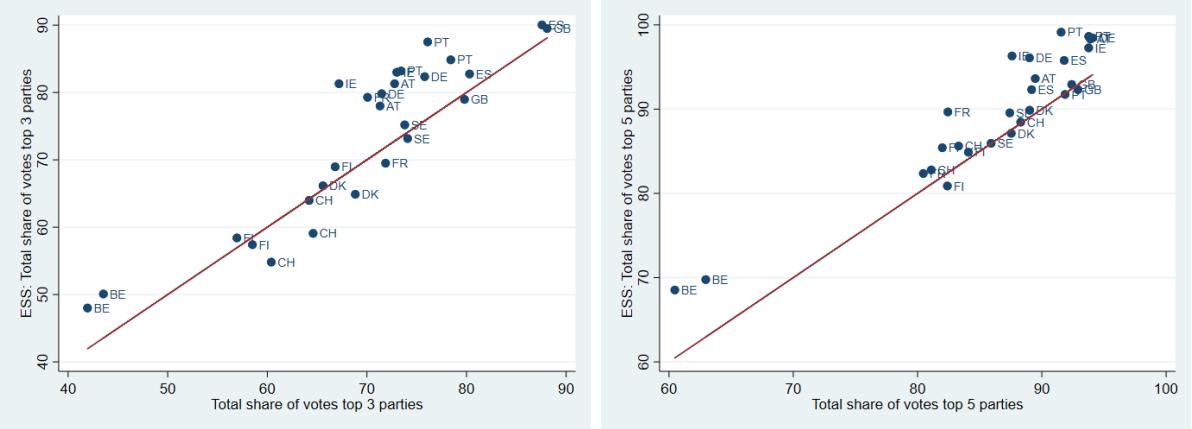
Voting Data		Obs.	Mean	Std. Dev.	Min	Max
Individual Characteristics	Age	48767	52.264	16.899	18	90
	Female	48767	0.502	0.500	0	1
	Tertiary	48767	0.356	0.478	0	1
	Tertiary (father)	40916	0.194	0.395	0	1
	Voting	78814	0.796	0.403	0	1
	Preferences Pro EU	48319	0.016	0.028	-0.242	0.080
	Preferences Pro National Way	48319	0.012	0.018	-0.061	0.116
	Nationalism (PCA, std)	48319	0	1	-1.607	8.113
Regional Characteristics	GDP per capita	261	31.644	11.749	13.267	69.901
	Unemployment rate (%)	261	9.632	5.491	0	30.1
	Tertiary rate (%)	261	30.69	8.038	11.2	55.1
	Share of Mig (%)	261	13.08	8.75	1.24	48.9
	Share of Mig (HS) (%)	261	3.31	3.1	0	19.1
	Share of Mig (LS) (%)	261	9.68	6.13	0.9	33.9
	Share of Mig (not EU) (%)	261	7.11	5.02	0.2	30.6
	Share of Mig (not EU, HS) (%)	261	1.69	1.62	0	13.4
	Share of Mig (not EU, LS) (%)	261	5.42	3.79	0	22.6
	Share of Mig (EU) (%)	261	5.88	5.52	0	34.4
	Share of Mig (EU, HS) (%)	261	1.62	1.86	0	11.7
	Share of Mig (EU, LS) (%)	261	4.26	3.88	0	24.5

Note: authors' calculations on ESS and ELFS data.

Table 3.2 describes the final sample used in the analysis. This includes 48,767 individuals who voted during the last election and are between 18 and 90 years old. The sample is balanced between males and females. Over the whole population, the average turnout is around 79%. Among the respondents, 35% have some tertiary education, and 19% have a father with tertiary education. On average, the party for which these individuals voted mentioned the EU in favourable way (the positive minus negative incidence is 1.6%) and also the national way of life (the net average positive incidence of mentions of national way of life is 1.2%). The first principal component of the preference for EU and preference for national life is the variable we call"Nationalism". There is a very strong negative correlation between EU preferences and national way of life preferences, so the first principal component is a linear combination of these two and it captures a very large fraction of common variance. We standardize the index so that a one-unit change of the index corresponds to one standard deviation. The lower part of Table 3.2 shows summary statistics of the variables measured

at the regional (NUTS2) level. The average population share of extra-EU immigrants in the considered sample (2007-2016) was equal to 7.1% of the population and it was larger than the population share of EU immigrants (5.9% of the population). Less-skilled immigrants account for about three fourths of total migrant population (they equal 9.7% of the population while high-skilled immigrants are 3.3% of total population respectively).

Figure 3.1: ESS vs. Actual Data on Voting



(a) ESS and actual votes - Top 3 parties

(b) ESS and actual votes - Top 5 parties

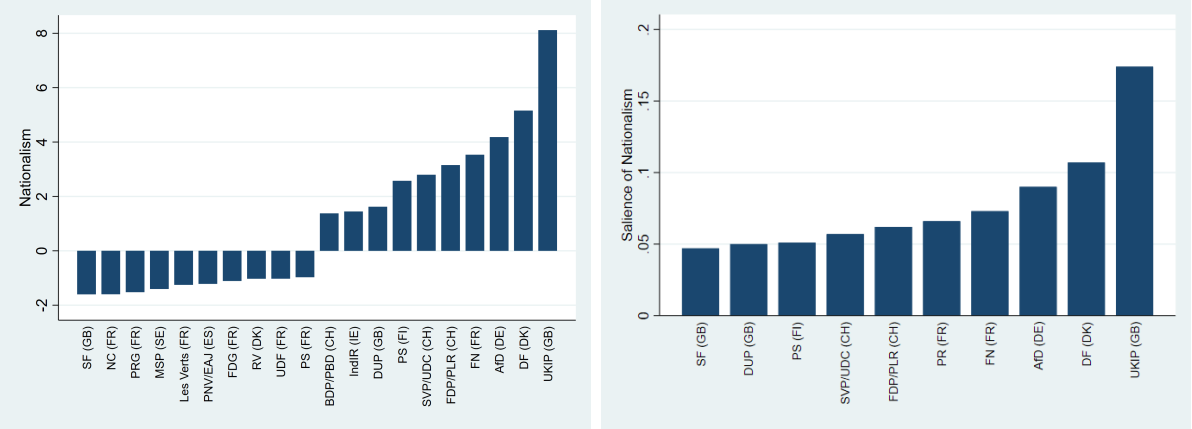
Note: authors' calculation on ESS and European Election Database (EDD). The figure plots the share of votes of top 3 parties (panel (a)) and top 5 parties (panel (b)) in terms of votes in each election available computed with the EDD (x-axis) and ESS (y-axis).

The data on individual votes obtained from the ESS is a sample. We consider it as representative of the aggregate vote of citizens at the national level. To check that there is no systematic bias, we compare the aggregate vote shares for the largest parties at the national level using our dataset, with the official figures reported in the elections. We use the European Election Database (EED), which is the most reliable source of information on vote shares, to measure the party vote shares at elections. Figure 3.1, panel (a) plots the total share of votes to the top three parties in each national election obtained using ESS data (vertical axis) against the same share obtained from the EED (horizontal axis). Panel (b) performs the same exercise, focusing on share of votes received in each election-country by the top five parties.[12] As there may be measurement error in the ESS, we would like to check if there is systematic under or over measurement of votes going to the top parties in the analyzed election. Figure 3.1 shows that country-election observations are close to

[12]Table A8 in the on-line Appendix shows the results in terms of differences between ESS and EED for each election for the top 5 parties. Considering the whole set of elections, the average difference between the two datasets is small (about 0.63% on average).

the 45 degree line, which would imply the exact same measure from the two sources. The departures are relatively small, sometimes positive and sometimes negative, and they show no specific pattern. This reassures us that the aggregate distribution of votes recorded in the ESS is representative of the official vote and the measurement error is not too large.

Figure 3.2: Salience and Preference for Nationalism across Parties



(a) Nationalism (top/bottom 10 parties) (b) Salience of Nationalism (top 10 parties)

Note: authors' calculations on Manifesto Project Database. The figure plots parties with the highest and lowest level of nationalism index (panel (a)) and parties with highest salience of nationalism (panel (b)) in their political manifesto. Nationalism in panel (a) is computed as the principal component over the net favorable mentions (measured as the difference between positive and negative mentions) of the European Community/European Union, and the National way of life in parties' manifestos. Salience of Nationalism in panel (b) is measured as the principal component over the total mentions (i.e. the sum of positive and negative ones) of the same topics in parties' manifestos.

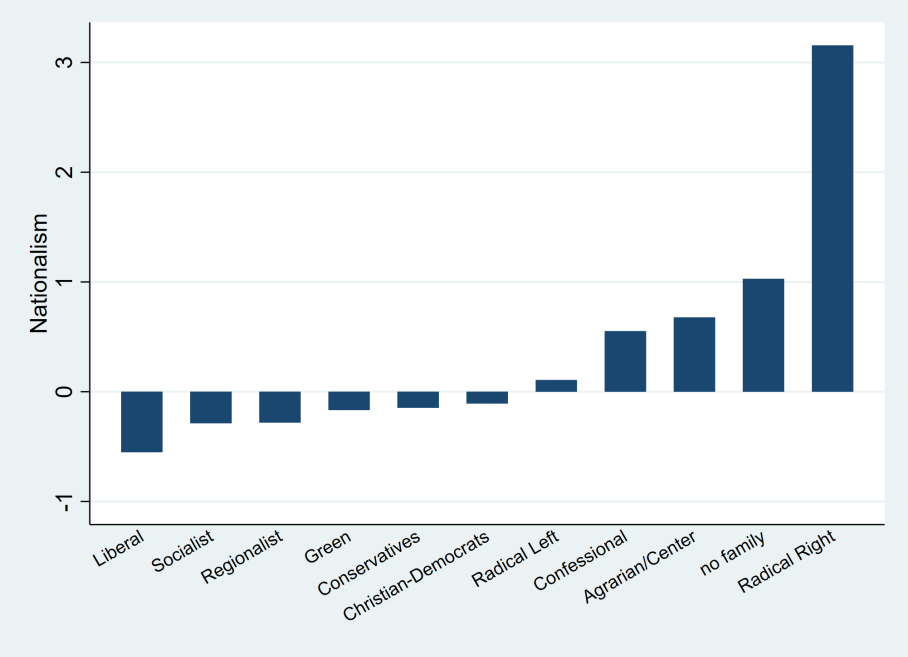
.

The next few figures describe the variation of the nationalism index across parties and regions. Figure 3.2 describes the range of variation across parties. The figure shows that some parties have a very strongly nationalistic platform and focus. Panel (a) of Figure 3.2 shows the value of the nationalism index for the top and bottom ten parties, according to that index, in Europe.[13] Parties with the most nationalist policy agendas are EU-skeptical, anti-immigrant, and occupy the right-wing of the political spectrum. The top one, the UK Independence Party, was one of the more aggressive supporters of Brexit, in the campaign preceding the referendum in which the UK voted to leave the European Union. The second one, the Danish People's Party, has brought an extreme nationalistic view in a country usually considered rather tolerant. The fourth, the Front National in France, had significant electoral success at the recent French presidential elections, with a very strong anti-immigrant

[13]The measure of preference for and salience of nationalism for each party in our sample is available in the on-line appendix in Table A13 and Figure A1.

message. Many of these parties are also in the top ten in terms of nationalism's salience (see panel (b)), namely how much of their manifesto they devote to issues of nationality and the EU relative to all other issues. At the other end of the spectrum, parties with the less nationalist agendas support a more active role for European institutions and take a more positive view of immigrants and more liberal immigration policies. They are usually left-wing parties (e.g., Nouveau Centre, Parti Radical de Gauche, Les Verts), center-left parties as the Moderate Coalition Party (MSP), as well as regional minority parties that support a European policy agenda to escape national centralism (e.g., "We Ourselves" in Northern Ireland; the Basque Nationalist Party in Spain).

Figure 3.3: Nationalism by Party Family



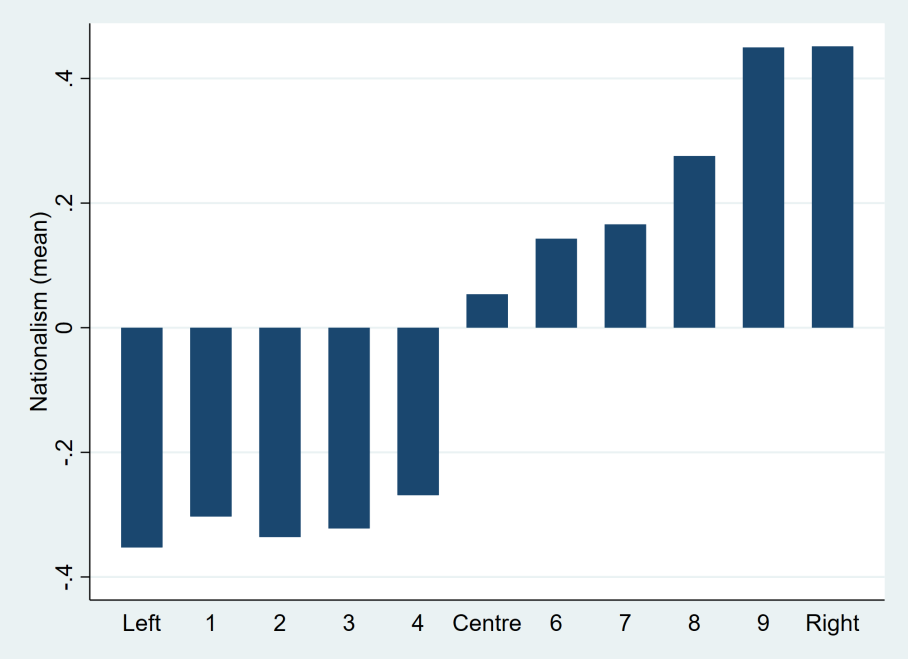
Note: authors' calculation on Manifesto Project Database and Chapel Hill Expert Survey Database. The figure plots the average level of nationalism index by parties' political families. Political families are identified by the Chapel Hill Expert Survey Database, using parties self-identification and belonging in the Europeans parties group.

Figure 3.3 shows the average value of the nationalism index when we aggregate parties into broad "political families" according to the *Chapel Hill Expert Survey* dataset. This classification aggregates parties into relatively coherent groups, based on their ideology and policies.[14] This picture provides an idea of where nationalism resides in European parties.

[14]The Chapel Hill Expert Survey asked two hundred political experts to classify parties using the Hix and Lord (1997) classification. They also use information about affiliation of national parties to international party associations to identify such political families. They identify eleven political families: Radical Right, Conservatives, Liberal, Christian-Democrats, Socialist, Radical Left, Green, Regionalist, Confessional,

We assign parties in the ESS to political families using parties' membership, affiliation or self-identification with a super-national group, as defined by the Chapel Hill Expert Survey dataset. Unsurprisingly, the nationalism index is highest among parties belonging to the "Radical Right" group and among those with high attachment to local territory and culture (e.g. Agrarian/Center and Confessional families). Nationalism is relatively low among parties belonging to the liberal and socialist traditions.

Figure 3.4: Nationalism by Left-Right Self-Declared Political Position



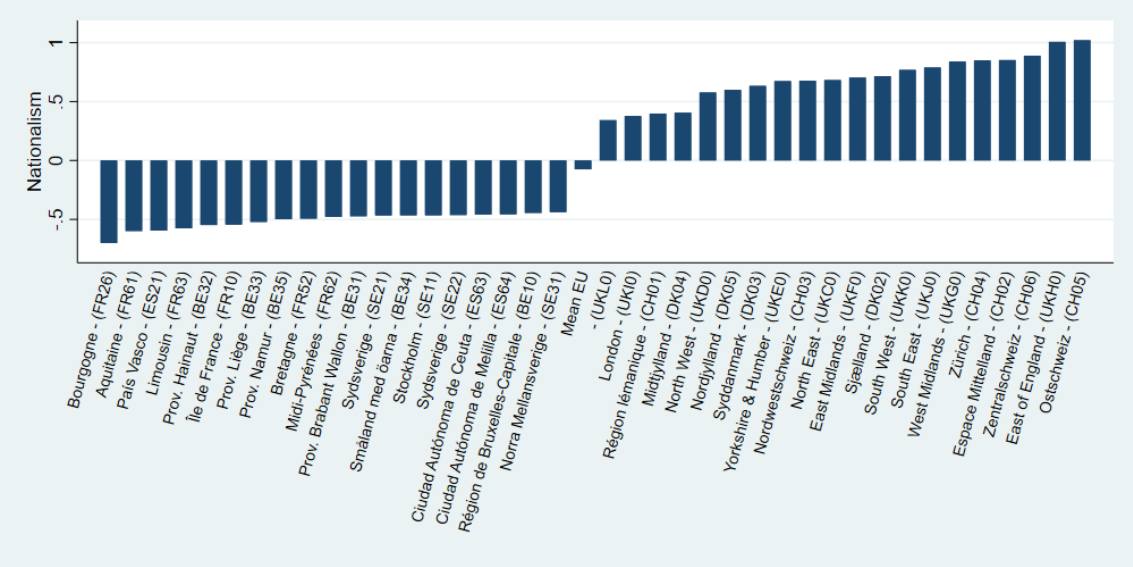
Note: authors' calculation on ESS and Manifesto Project Database. The figure shows the average level of nationalism index associated with each individual self-declared political position on the axis Left-Right.

Figure 3.4 shows the average nationalism associated with a position in the political spectrum that ranges from 0 (extreme left) to 10 (extreme right). Moving from from left to right across the political spectrum, and capturing with each bar a decile of the voters, the intensity of individual preferences for nationalism increases almost monotonically moving from left to right along the political spectrum. Individuals who self-identify as "left-leaning" tend to vote for anti-nationalist policy agendas. To the contrary, "right-leaning" individuals vote for parties with more pro-nationalist agendas.

Figure 3.5 shows the range of variation in the average nationalism index across European regions. We only report the top and the bottom twenty European regions in terms of average nationalism. Regions where people voted for the least nationalist political agenda are in

Agrarian and a residual "No family" group.

Figure 3.5: Average Nationalism across Regions - Top/Bottom 20



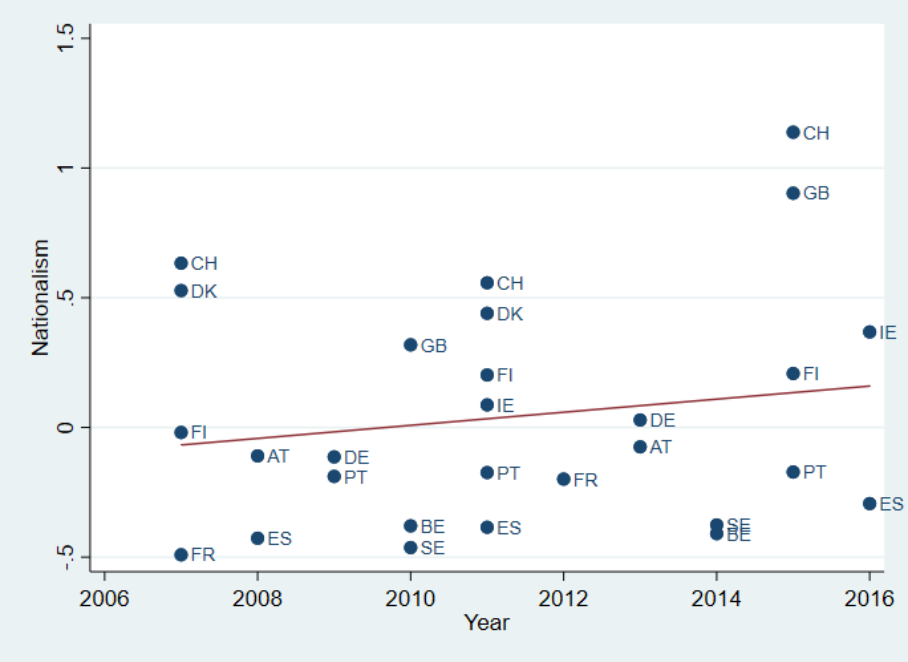
Note: authors' calculations on ESS and Manifesto Project Database. The figure plots the highest and lowest average level of nationalism index at regional level in our sample.

Spain, France, Belgium and, to a lesser extent, Sweden. More nationalist voting comes from regions in countries that are less-engaged by European economic and political integration, such as Switzerland, UK, and Denmark. Voters in these regions privilege a nationalist policy agenda. This figure shows a cross-section in the average levels of the nationalism index. In our empirical analysis we consider, instead, changes in the index within regions over time as our variation of interest, and we try to explain it using changes in the immigrants share.

In order to show that the vote for nationalist parties, and hence the average nationalism index, has increased over time for most countries in our sample, Figure 3.6 plots nationalism levels for countries and years in which elections were held. We see a clear upward trend in average nationalism over time during the ten years considered. Even when looking at individual countries, both those with initially-high (e.g. Switzerland and the UK) and initially-low (Spain and Sweden) levels of nationalism, we observe a generalized increase in the index. Figure 3.6 is produced by keeping the nationalism index of each party fixed at its initial level and measuring the average by weighting each party with the regional share of voters for the party in that election.

In the statistics shown so far we have fixed the nationalism of a party to its average level over the period and we have shown regional and national evolution due to change in vote. However with our data we can also track the changes in the nationalism index for a party over time. Are average changes in country attitudes–in terms of nationalism–due to change

Figure 3.6: Country-specific Nationalism Over Time



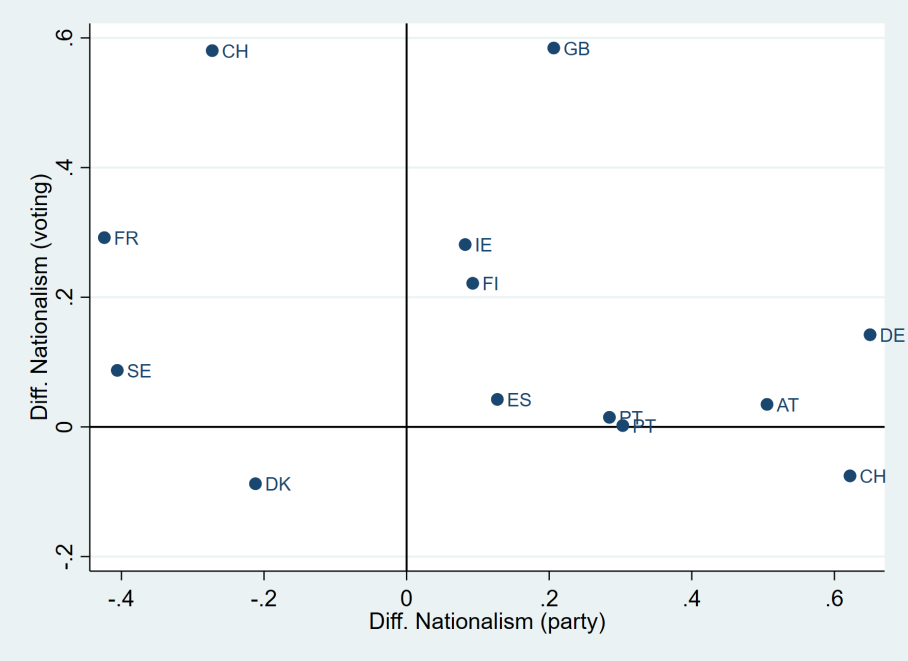
Note: authors' calculations on ESS and Manifesto Project Database. The figure plots the average country level of nationalism after each election event in our analysis.

in votes or in changes of party platforms? In Figure 3.7 we show for each country the average change in the level of nationalism between electoral events, when this is measured keeping party position fixed and looking at the change in voting (y-axis), and when we keep the voting share fixed and consider changes in party platforms (x-axis). Figure 3.7 shows mostly increases in the nationalism index both at the voting and party platform level. And in the majority of the cases (8 countries out of 13) both changes are positive. The main focus of our analysis is the change in voting behavior of natives in response to immigrants, hence we will leverage mainly the variation shown in the vertical axis of Figure 3.7.[15]

Figure 3.8 shows the panel variation of the main explanatory variable, namely the share of high- (panel (a)) and less- (panel (b)) educated migrants in the total population, following the same timing of the available elections. Overall, we notice a positive upward trend, which is steeper for highly-educated immigrants. Notice, however that less-educated immigrant share of the population was almost double that of highly-educated. Comparing Figure 3.6 and Figure 3.8, we can point out a common upward trend for the two variables, which is

[15]In our baseline specification, we consider the political platform of the voted-for parties as fixed at its average level. We then perform sensitivity analyses considering the platform at its initial or final position (see Table A1 in the Appendix). We also analyze the change in the political platforms of parties in response to migration, showing that this is qualitatively similar to the change in voters' responses. These results are available upon request.

Figure 3.7: Nationalism: Voters or Parties?



Note: authors' calculations on ESS and Manifesto Project Database. The figure decomposes the country-specific variation in Nationalist voting between electoral events in variation due to individual voting behavior (given parties political platforms) and variation due to parties political agendas (given individual votes). The x-axis shows the variation driven by changes in parties political agenda, while y-axis shows the variation driven by a variation of individuals voting behavior.

Figure 3.8: Share of High-skilled and Low-skilled Immigrants Over Time

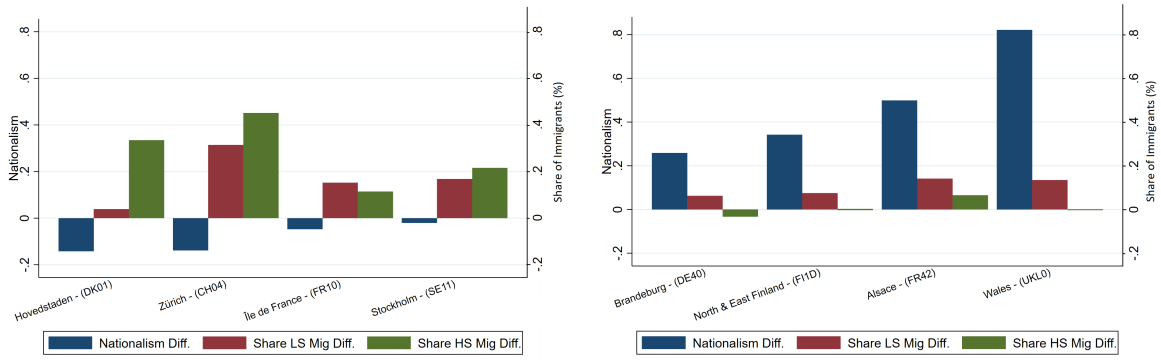
(a) Share of high-skilled immigrants

(b) Share of low-skilled immigrants

Note: authors' calculations on EULFS data. The figure plots the average share of immigrants over total population in 2005 at regional level by country and education.

interesting but not causal, and which we will control for with time fixed effects.

Figure 3.9: Variation of Nationalism and Skill-Specific Immigration



(a) Reduction of Nationalism

(b) Increase of Nationalism

Note: authors' calculations on ESS, Manifesto Project Database and EULFS. The figure plots the variation of NUTS2 average level of Nationalism through voting and the skill-specific variation of the share of immigrants for regions in the sample experiencing the largest decrease in Nationalism during 2007-2016 (panel a) and regions experiencing the largest increase in Nationalism (panel b).

In order to provide an idea of the panel variation across regions, we show in Figure 3.9 four regions characterized by large decreases in the nationalism index between 2007 and 2016 in panel (a), and four regions with large increases in the nationalism index in panel (b), both represented by the blue bars. In the former group we find regions with large metropolitan areas such as Copenhagen, Zurich, Paris, and Stockholm. These are areas characterized by relatively large inflows of migrants, and in most cases the net inflow of high-skilled (green bar) prevails on that of less-skilled (red bar). To the contrary the regions with large nationalism increase are less densely populated areas, where the inflow of immigrants is generally smaller and where the inflow of less-skilled is larger than the inflow of high-skilled. This picture is consistent with the idea that highly-educated immigrants move to dense urban areas and their inflow may contribute to reduce the level of nationalism. Conversely, few–but mainly less-skilled–immigrants may contribute to steer less populated areas towards nationalism.

3.3 Empirical Specification and Identification

In this section we present our basic empirical approach to investigate the impact of immigration on the nationalism of EU citizens as revealed by their vote. As key explanatory variable we include, in the basic specification, the adult population share of high- and less-educated immigrants. As our goal is to focus on how immigrants may change the vote of EU

citizens, we only include natives' voting as an outcome. This eliminates a potential direct effect of naturalized immigrants on voting which could derive from their different political preferences. We define as $Nation_{i,r,e}$ the normalized value of the nationalism index (first principal component of favorable mentions toward the EU and national way of life) for the party voted for by native citizen i in region r (NUTS2) in election e. In our basic specification, the measure of nationalism of a party is taken as fixed the first time the party appears in the sample. We also control for additional individual and regional characteristics which may affect individual preferences. The basic specification is as follows:

$$Nation_{i,r,e} = \alpha + \beta_H m^H_{r,e} + \beta_L m^L_{r,e} + \zeta \mathbf{X}_{i,r,e} + \gamma \mathbf{\Gamma}_{r,e} + \theta_e + \eta_r + \epsilon_{i,r,e}. \tag{3.2}$$

The two explanatory variables of interest are $m^H_{r,e}$ and $m^L_{r,e}$. They represent the number of migrants with tertiary (H) schooling[16] and the number with secondary or less (L) schooling, respectively, as share of the total 2005 population in region r in election e. We standardize for total adult population in 2005, so that the variation of that ratio is only driven by changes in the immigrant population and not by native-population changes.[17] The vector $\mathbf{X}_{i,r,e}$ includes a set of individual controls capturing characteristics of voter i. They include age, gender and whether the individual is tertiary educated. The vector $\mathbf{\Gamma}_{r,e}$ contains a set of aggregate characteristics of region r including the logarithm of the average GDP per capita, the unemployment rate and the percentage of tertiary-educated individuals in the regional population. We control for time-invariant regional factors, such as institutions, local culture and geographic location, with the inclusion of regional fixed effects η_r. Aggregate changes over time, between elections are captured by the election-year fixed effects θ_e.

Estimating equation (3.2) using OLS produces a measure of the partial correlation between the high- and less-skilled immigrant share and natives' voting preferences. The estimates of coefficients β_H and β_L capture such correlations. However, it is likely that unobserved regional characteristics correlated with the error term $\epsilon_{i,r,e}$ may also affect immigrants' location decision. For instance, a better economic outlook in a region or specific regional policies may attract immigrants and reduce nationalism in the region. If this is case, the OLS estimation will produce a biased estimate of the causal effect of immigration on votes. To mitigate this potential bias, we use an instrumental variable approach based on a variation of the shift-share methodology (see Card, 2001, Mayda *et al.*, 2018) which is widely used

[16]Specifically, we define highly-educated immigrants as those who have completed at least one year of tertiary education which corresponds to level 5 or above of the ISCED classification.

[17]In the empirical analysis we will also consider separately the number of immigrants originating from other EU28 countries, and those originating from outside of EU28.

in the immigration literature. The intuition for it is to use past location of immigrants by country-of-origin across regions as predictor of subsequent inflows of immigrants. The assumption is that past settlements of immigrants from specific countries are uncorrelated (or less correlated) with current factors that may affect the political preferences of individuals. Then, the net aggregate inflows of immigrants from each country of origin in the period 2007-2016, likely driven by aggregate push factors or by changes in immigration laws (such as the EU enlargement), are allocated to EU regions according to the early distribution of immigrants from those countries. Hence, such a shift-share or "enclave" instrument generates variation of regional migration over time due to the interaction between previously established immigrants' settlement and aggregate national flows. Both of those components are likely to be less correlated with region-specific economic and political changes during the period, relative to actual migrant flows. This should reduce the severity of the omitted variable bias. We will also test the correlation of the instrument with the pre-period (2000-2007) economic and demographic regional trends. This test will ascertain whether the predicted inflows are correlated with pre-existing economic and demographic trends which, if persistent and if they affect voting, may drive a potential omitted variables bias. These checks are an important step in reassuring the reader of the validity of the instrument and reducing the worries raised in recent criticism of the shift-share IV strategy (see Jaeger *et al.*, 2018 and Goldsmith-Pinkham *et al.*, 2018).

As we want to capture skill-specific effects of immigration, we construct two different instrumental variables. They use the aggregate variation of immigrant flows by country-of-origin for each separate education group (high- and less-educated), and we interact those with the same initial (2005) distribution of immigrants, by origin, across regions.[18] Namely we construct the two instruments as follows. Define the initial presence of immigrants from origin country c in year 2005 in each NUTS2 region r as the share of their total population in 12 European countries in 2005 as follows:

$$sh_{c,r,2005} = \frac{M_{c,r,2005}}{\sum\limits_{r \in EU_{12}} M_{c,r,2005}}, \tag{3.3}$$

where $M_{c,r,2005}$ is the stock of immigrants born in country c and residing in region r in year 2005, while the denominator is the total stock of immigrants born in country c and residing in all countries of our sample as of 2005. Year 2005 is the earliest year available in the EULFS for which we can construct such a regional distribution. The EULFS, in its harmonized codification scheme, grouped the origin countries into 15 groups that we use as

[18] This is also done in Mayda *et al.*, 2018.

"countries" of origin.[19] Then, we compute the total stocks of immigrants resident in the EU 12 countries, born in country-of-origin c in each year $t \in \{2007, 2008, ..., 2016\}$ separately for each skill levels s equal to college educated, H, or non-college educated, L. Namely we compute:

$$TM^s_{c,t} = \sum_{r \in EU_{12}} M^s_{c,r,t} \text{ for } s = H, L. \tag{3.4}$$

We can then predict the total number (stock) of immigrants of skill s, origin c in region r and year t as the interaction (product) of expression (3.3) and (3.4), as follows:

$$\widehat{M}^s_{c,r,t} = sh_{c,r,2005} * TM^s_{c,t}.$$

We then sum across all countries of origin to obtain the total predicted number of immigrants of each education type, and standardize for the total adult population in region r in year 2005, to have the predicted skilled and less-skilled immigrant as share of the initial population in region r and year y as follows:

$$\widehat{m}^s_{r,t} = \frac{\sum_c \widehat{M}^s_{c,r,t}}{Pop_{r,2005}}.$$

These imputed shares are used as instrumental variables for the actual immigrants of each skill level H and L as a share of the 2005 population. The predictive power of this IV, for each skill group, will depend on the tendency of that group to locate where previous immigrants from the same origin did. Notice that the 2005 distribution of origins across regions is not skill specific, as to avoid correlation with regional features that would select high- or less-skilled workers. Clearly, if there is a very strong correlation between skill and origins' distribution across destination regions, this can still generate a correlation between the initial immigrant distribution with local attributes and the evolution of nationalism. Hence the relative variation of the two instruments is generated by the differential aggregate

[19] The fifteen birthplace regions identified by the EULFS data are the following: Natives, EU15, New European member state that joined EU in 2004, New European member state that joined EU in 2007/2013, EFTA countries, Other Europe, North Africa, Other Africa, Near and Middle East, East Asia, South and South East Asia, North America, Central America and Caribbean, South America and Australia-Oceania. Additional information on data and more details on the construction of the IV is available in 3.9.2.

flows of immigrants of each skill-level across countries-of-origin. Large differences in the education levels of immigrants across areas-of-origin, as well as large differences in the growth of each immigrant group, generate the variation needed to identify the two effects, of less- and more-educated migrants, separately.

Instrumental variables based on the shift-share methodology have been used before to explore the relation between immigration and voting (see e.g. Otto and Steinhardt, 2014; Mendez and Cutillas, 2014; Brunner and Kuhn, 2018; Halla *et al.*, 2017; Barone *et al.*, 2016). They have also been subject to a fair amount of discussion and criticism (e.g. Jaeger *et al.*, 2018 and Goldsmith-Pinkham *et al.*, 2018). Persistent local conditions that influence immigrants' location decision and natives' voting preferences can threaten this identification strategy by allowing omitted variable bias that would affect the OLS estimated coefficient up or down with respect to the causal effect of immigration on the vote. To reduce those concerns, we will check the correlation between the instrument and pre-2005 economic and demographic regional trends. Moreover, in our case two sources of variation in the flows of immigrants in the post-2007 period were important and orthogonal to the presence of previous immigrants and to the local political sentiment. First, access to the EU and the implementation of the Schengen Agreement between 2004 and 2006 for nine new EU member countries [20] produced an increased inflow of immigrants from these new EU countries to the countries of Western Europe. Second, the widespread economic and financial crisis of 2007-2009 increased emigration, especially of highly-skilled individuals, from the countries of Southern and Eastern Europe, generating important exogenous variation of these flows in receiving regions. Both shocks were large and independent of previous regional conditions and likely not correlated with past shocks. Hence, if these aggregate shocks drove a large part of the variation in national migrations, it would suggest that the IV variation was quite independent from the unobserved regional shocks. Table 3.3 shows the coefficients obtained from regressing the 2007-2016 change in the IV, across regions, on pre-2005 growth of GDP per capita, employment rate, population, share of tertiary-educated and life expectancy. Specifically, the dependent variable is the imputed growth of immigrants over the period 2007-2014, for all immigrants (column 1), for highly-educated immigrants (column 2), and for less-educated immigrants (column 3). The independent variable is the 2000-2007 growth of GDP per capita (row 1), of employment rate (row 2), of population (row 3), of the share of tertiary educated (row 4), and the life expectancy growth rate (row 5). Most of the correlations are not significant at any reasonable confidence level, revealing that the predicted change in im-

[20] Signed on the 16th of April 2003 by Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia and Slovenia, the Schengen Agreement became operative for those countries on December 2007.

migrants (high- or less-skilled) is not associated with pre-existing regional trends. Only one of the fifteen coefficients, between predicted growth of high-skilled from 2007-2014 and GDP per capita growth 2000-2007, is significant at the 10% level. Overall, the small value and low significance of these coefficient estimates suggest that the shift-share-predicted inflows of immigrants are not correlated with pre-existing regional trends, and this is consistent with the validity of the instrument.

Table 3.3: Correlation between 2000-2007 Growth of Economic Indicators and Post-2006 Predicted Migrants Growth

	(1)	(2)	(3)
Predicted Mig growth	*All*	*HS*	*LS*
GDP per capita growth rate	-0.09	-0.19*	-0.05
	(0.05)	(0.09)	(0.04)
Observations	96	96	96
Adj. R-Square	0.06	0.16	0.01
Employment growth rate	-0.15	-0.29	-0.08
	(0.11)	(0.18)	(0.10)
Observations	99	99	99
Adj. R-Square	0.04	0.09	0.01
Population growth rate	-0.13	-0.28	-0.08
	(0.11)	(0.18)	(0.09)
Observations	109	109	109
Adj. R-Square	0.02	0.06	0.00
Tertiary growth rate	-0.04	-0.10	-0.02
	(0.03)	(0.06)	(0.03)
Observations	99	99	99
Adj. R-Square	0.02	0.10	0.00
Life Exp. growth rate	0.51	0.80	0.31
	(0.97)	(1.00)	(1.09)
Observations	93	93	93
Adj. R-Square	-0.00	-0.00	-0.01

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Standard errors are clustered at country level. * p<0.1, ** p<0.05, *** p<0.01. The table shows the predicted coefficients regressing the growth rate of regional macro indicators between 2000 and 2007 on the regional predicted migration growth with our shift-share strategy over the period 2007 and 2014.

3.4 Baseline Results

Table 3.4 presents our baseline results. The reported coefficients capture the partial correlation between immigrants as a share of the initial population in a region and the intensity of nationalism of voters in that region. The top panel in the table shows the estimates using all immigrants as explanatory variable. The lower panel shows results when only considering migrants from non-EU origins.[21] For each set of estimates, we estimated one specification using OLS (columns (1), (3), (5), and (7)) and one using IV (columns (2), (4), (6), and (8)). In columns (1) and (2), the explanatory variable is the total share of immigrants relative to the initial total population of residents (as of year 2005). We keep the population in the denominator fixed to 2005, so that it serves as a standardization and is not affected by potential changes in native populations over the same period. In columns (3) and (4) the explanatory variable is the high-skilled-immigrant share of the 2005 population, and in columns (5) and (6) it is the less-skilled-immigrant share of the initial population. In columns (7) and (8) we include shares of both high-skilled and less-skilled in the same regression. All specifications include regional fixed effects, election fixed effects, plus the following regional and individual controls: age, gender and tertiary education dummy at individual level, the logarithm of GDP per capita, the percentage of tertiary-educated individuals, and the unemployment rate at the regional level.

The F-stat of the first stage (reported in the table for all IV estimates) shows that the shift-share instrument has significant power. Even when the two shares (high- and less-skilled) are included together as explanatory variables and instrumented, the F-stat remains above 10, indicating enough differential variation in the inflow of high- and less-skilled immigrants across origins. The baseline estimates in the top panel of Table 3.4 show that, on average, a higher share of immigrants does not have a significant impact on the intensity of nationalism in voters.[22] However, when separating the effect of less- and high-skilled immigrants, we see a clear pattern and significant difference. An increase in the share of less-skilled immigrants in a region is associated with increase in nationalism. Conversely, an increase in the share of high-skilled immigrants is associated with a decrease in nationalism.[23] The size of the

[21] When considering the share of EU immigrants only, the instrument has very low power, hence we do not report those results.

[22] Our population of voters includes individuals that were born in the country-of-destination; i.e., also second- and higher-order generations of migrants.

[23] This result is consistent with the main finding of Mayda et al (2018) on US elections. They find that an increase in skilled immigrants in the US reduces the percentage of votes to Republicans, while an increase in less-skilled immigrants increases the Republican share of votes. Moreover the results are confirmed when we associate with each party the average level of nationalism or the last available measure of nationalism, as Table A1 shows.

Table 3.4: Nationalism Intensity and Immigrant Share

	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
Time	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016
All Migrants								
Share	-0.01	0.01						
	(0.01)	(0.05)						
Share HS			-0.06*	-0.14*			-0.07**	-0.17**
			(0.03)	(0.07)			(0.03)	(0.07)
Share LS					0.01	0.05	0.03*	0.08**
					(0.02)	(0.05)	(0.02)	(0.04)
Observations	48303	48303	48303	48303	48303	48303	48303	48303
K-P rk Wald F-stat		45.42		32.24		38.72		14.96
Adj. R-Square	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
Not EU Migrants								
Share No EU	-0.01	0.04						
	(0.02)	(0.05)						
Share No EU HS			-0.09*	-0.26			-0.10*	-0.43*
			(0.05)	(0.20)			(0.05)	(0.22)
Share No EU LS					0.00	0.07	0.02	0.14***
					(0.02)	(0.05)	(0.02)	(0.04)
Observations	48303	48303	48303	48303	48303	48303	48303	48303
K-P rk Wald F-stat		40.48		39.42		44.70		17.00
Adj. R-Square	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
NUTS2 f.e.	✓	✓	✓	✓	✓	✓	✓	✓
Year f.e.	✓	✓	✓	✓	✓	✓	✓	✓
NUTS2 Controls	✓	✓	✓	✓	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Standard errors are clustered at NUTS2 level. * p<0.1, ** p<0.05, *** p<0.01. As individual controls age, gender and education dummy are included. As NUTS2 controls we include log GDP per capita, unemployment rate and share of tertiary educated individuals. The dependent variable is our measure of nationalism, standardized to have zero mean and unity standard deviation.

coefficients is larger using IV than using OLS estimations. This suggests IV may correct for omitted factors that produce an attenuation bias in OLS estimates.[24] Estimates relative to the share of non-EU immigrants, shown in the lower panel of Table 3.4, confirm the findings. The positive effect of less-skilled migrant shares on nationalist voting is even larger when

[24]e.g. less-skilled immigrants may decide not to go to European regions characterized by a high intensity of nationalist voting. Similarly, omitted economic or social factors at the regional level may determine both low immigration rates and pro-nationalist voting. We discuss the role of local regional conditions more in detail in Table 3.7 below.

considering only immigrants from extra EU countries (see IV estimates in columns (6) and (8)). The anti-nationalist voting effect of skilled migration from non-EU migrants is larger in size, but less precisely estimated compared to the full set of migrants (see columns (4) and (8)).

We can evaluate the economic magnitude of these effects by using the estimated marginal effect of migration shares, together with the corresponding standard deviations reported in Table 3.2. Taken at face value, the IV estimates in column (8) in the top panel of Table 3.4 indicate that a one standard deviation increase in the share of less-skilled migrants (6.1 percentage points) shifts voting of individuals toward a more nationalist agenda by $0.49(= 6.1 * 0.08)$ standard deviations of the nationalism indicator. This is a sizable effect, almost equivalent to the distance in nationalism between the Radical Left family (0.05) and a Confessional (0.55) family. The magnitude of the negative effect of a one standard deviation increase in the share of high-skilled migrants on nationalist voting is equal to $-0.53(= 3.1 * (-0.17))$ standard deviations. It is roughly equivalent to shifting voting of individuals from the agenda of a Radical Left party, guided by economic concerns regarding the European integration process, to the views of most Socialist or Liberal parties in Europe. While the coefficient for high-skilled immigrants is larger then for less-skilled immigrants, their variation across regions is smaller, hence a similar effect is produced in terms of magnitude.

These results are obtained considering the average nationalism of parties over 2006-2016. In Table A1 in the Appendix, we report very similar findings when fixing party nationalism at the initial or final value observed in the sample. Table A2 analyzes whether the electoral vote response to immigrants also reveals a change in the salience of nationalism in the party agenda. The estimates suggest that the effect of immigration on the index of salience of nationalism in voted-for parties is similar to the effect on the intensity of nationalism itself. Less- and high-skilled migration have opposite effects on salience. A higher share of less-skilled migrants makes nationalism a more salient topic among voted parties. Conversely, a higher share of high-skilled migrants reduces the salience of nationalism among voted parties. Signs of the coefficients are stable across specifications. As in the previous results, coefficients are larger using the IV strategy relative to OLS estimates. Effects are very similar when we concentrate on non-EU migrants

Table 3.5 shows the robustness of the main coefficients as estimated in Table 3.4, column (8), top panel, to the inclusion/exclusion of individual and regional controls, using the IV estimation strategy. In column (1), we drop from the specification the set of individual and regional controls and we keep only the regional and election fixed effects. The coefficients are estimated less precisely, as some omitted variable bias may arise, but the signs are consistent

Table 3.5: Nationalism and Immigrant Share - Including Controls

	(1)	(2)	(3)	(4)
	IV	IV	IV	IV
Time	2007-2016	2007-2016	2007-2016	2007-2016
Share HS	-0.14	-0.13	-0.18**	-0.17**
	(0.09)	(0.09)	(0.07)	(0.07)
Share LS	0.19***	0.19***	0.07**	0.08**
	(0.06)	(0.06)	(0.04)	(0.04)
Female (*ind*)		-0.08***		-0.08***
		(0.02)		(0.02)
Age (*ind*)		-0.00		-0.00
		(0.00)		(0.00)
Tertiary (*ind*)		-0.19***		-0.19***
		(0.03)		(0.03)
ln(*GDP cap*) (*NUTS*2)			0.49	0.53
			(0.42)	(0.42)
Tertiary (*NUTS*2)			0.06**	0.06**
			(0.02)	(0.02)
Unemp. rate (*NUTS*2)			-0.00	-0.00
			(0.01)	(0.01)
Observations	48303	48303	48303	48303
K-P rk Wald F-stat	10.73	10.74	14.95	14.96
Adj. R-Square	0.12	0.12	0.12	0.13
NUTS2 f.e.	✓	✓	✓	✓
Year f.e.	✓	✓	✓	✓

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Standard errors are clustered at NUTS2 level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is our standardized measure of nationalism.

with those in Table 3.4, column (8), top panel. In column (2), we include a baseline set of individual characteristics i.e. a dummy for female/male, a dummy for tertiary education, and linear age. In column (3), we include the set of regional controls (i.e. per capita GDP (in logarithm terms), regional unemployment rate, and share of residents with tertiary education) without individual controls. Finally, in column (4) we include the complete set of

individual and regional controls, so as to obtain our baseline specification. Our main results remain reasonably stable across specifications. Among individual controls, being female and having tertiary education are associated with voting for parties with a significantly less-nationalistic agenda compared to being male or less-educated, respectively. Age does not have any significant correlation with the intensity of nationalist vote. Among regional controls, the positive significant coefficient on the share of residents with tertiary education is somewhat surprising and counter to the individual effect of education on nationalism intensity. Other regional controls do not seem to much affect the nationalism intensity of the vote, nor to affect the coefficient on immigrant share.

Another concern related to our baseline results is that individuals reveal their political preferences for policy platforms only indirectly, by voting for parties whose political agenda cover multiple issues. For this reason we investigate to what extent the mechanisms highlighted by our baseline analysis (and captured there by a constructed index) translate into actual voting for specific party groups. Figure 3.3 shows that Liberal and Radical-Right party families stand at opposite ends of the nationalist political spectrum. The former group presents a policy agenda which favors immigration, being open to economic, political and cultural European integration. The latter group has often strong anti-immigration views, and rejects EU integration on the basis of sovereignty arguments and cultural claims. We thus estimate a linear probability model to analyze how migration affects voting shares for parties belonging to each group. Results in Table 3.6 suggest that less-skilled immigration encourages voting for radical-right parties and has no effect on voting for liberal parties. Similarly, high-skilled immigration has no effect on voting for liberal parties, while reducing voting for radical-right parties. These effects are consistent with previous results and describe shifts toward opposite ends of the political spectrum in response to high and less-skilled immigrants.[25]

[25]In online 3.9.7 we further explore the effect of voting for radical-right and liberal parties by voters' education (Table A17). Results show that the effects of migration on the propensity to vote for radical-right and liberal parties are mainly driven by non-tertiary-educated voters.

Table 3.6: Voting for Radical Right - Liberal Parties and Immigrant Share

	(1)	(2)	(3)	(4)
	IV	IV	IV	IV
Time	2007-2016	2007-2016	2007-2016	2007-2016
Party Family	*Radical Right*		*Liberal*	
Share	0.00		0.01	
	(0.01)		(0.01)	
Share HS		-0.03***		0.04
		(0.01)		(0.03)
Share LS		0.01***		0.00
		(0.01)		(0.01)
Observations	45771	45771	45771	45771
K-P rk Wald F-stat	39.76	6.89	39.76	6.89
Adj. R-Square	0.10	0.10	0.12	0.11
NUTS2 f.e.	✓	✓	✓	✓
Year f.e.	✓	✓	✓	✓
NUTS2 Controls	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓

Note: authors' calculations on ESS, EULFS, Manifesto Project Database, Chapell Hill Database and Eurostat data. Standard errors are clustered at NUTS2 level. * p<0.1, ** p<0.05, *** p<0.01. As individual controls age, gender, education dummy are included. As NUTS2 controls we include the log of GDP per capita, unemployment rate and share of tertiary educated individuals. The dependent variable is a dummy that takes value of 1 if voter for a party that belongs to the Radical Right political family (col. (1)-(2)) or to the Liberal political family (col. (3)-(4)). The definition of the political family of parties is determined by the Chapel Hill Expert Survey dataset.

3.5 Heterogeneity in Individual and Regional Voting Responses

After analyzing the average effect of regional immigration on nationalism, we explore two important sources of potential heterogeneity in these effects. First, individuals may differently update their perceptions in response to the inflow of immigrants, depending on the perceptions and priors which are likely to vary with education, gender, age and area of resi-

dence. Second, we analyze whether effects of immigration vary depending on region specific characteristics. Perceptions of migrants' contributions to the local economy and society may be affected by some local features, such as the incidence of crime and generosity in the provision of public services such as schools. Previous studies have found that local provision of public goods (see Otto and Steinhardt, 2014, Becker and Fetzer, 2017), and children's education (see Halla *et al.*, 2017) can be important channels influencing the attitudes of natives toward immigration[26]. In the next two sub-sections we analyze how heterogeneity affects the response of nationalism to immigrants.

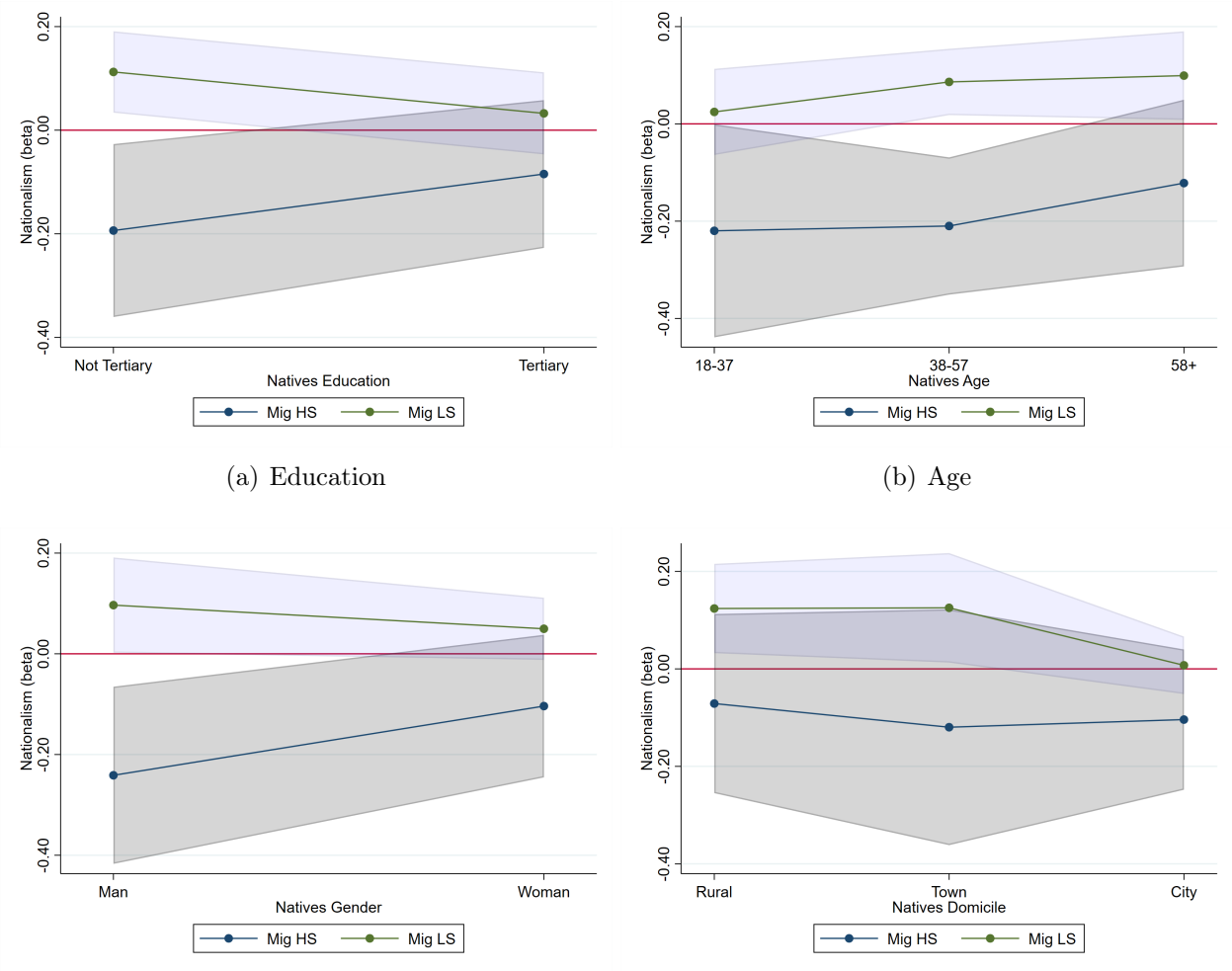
3.5.1 Estimates of Differential Individual Effects

To allow for heterogeneous effects of migration across voters' individual characteristics, we run regressions on separate sub-samples of citizens by their education levels, separating tertiary- and non-tertiary-educated by age (grouping 18-37, 38-57, and 58+ together), by gender (male and female), and by place of residence (separating into rural areas, cities, and small towns). We estimate different coefficients of the impact of immigrants on nationalism associated with each group. We summarize these coefficients in figures rather than tables. Figure 3.10 plots the estimated effect of high-skilled migration (Mig HS, blue line), and less-skilled migration (Mig LS, green line) for each of these subgroups. Shaded areas denote the 95% confidence intervals around the estimated value.

Panel (a) reports results by education level and shows a remarkable difference in the response of more- and less-educated natives to high- and less-educated immigrants. Interestingly, less-educated citizens are pushed towards nationalism significantly more than more-educated citizens, in response to less-skilled immigration (top green line). At the same time, they are *also* pushed away from nationalism more intensely than highly-educated citizens in response to high-skilled immigrants. A one standard deviation increase in the share of less-skilled migrants raises nationalist voting by citizens without tertiary education by $0.67(= 6.1 * 0.11)$ standard deviations of the nationalism indicator. This is a significant effect, comparable to the distance in nationalism between a Liberal party (-0.55) and a Confessional party (0.07). Symmetrically, less-educated natives are also rather responsive to high-skilled migration. A one standard deviation increase in high-skilled migration reduces nationalist voting by $-0.59(= 3.1 * (-0.19))$ standard deviations of the nationalism indicator. Overall, education of citizens seems to reduce the responsiveness of voting behavior to the migration phenomenon. This may consistent with a Bayesian updating of beliefs: less-skilled individuals may know

[26]Mendez and Cutillas (2014) do not find a specific effect of exposure to crime on attitudes toward immigrants

Figure 3.10: Nationalism Response to Immigrant Share and Natives' Individual Characteristics



(a) Education

(b) Age

(c) Gender

(d) Domicile

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. The graphs plots the coefficients of the share of migrants HS and LS estimated on different subsamples by individual characteristics: education (a), age (b), gender (c) and domicile (d). All the coefficients are estimated with IV estimations. The shadowed area represent the 95% interval of confidence. All the regressions includes individual and regional controls, NUTS2 and year fixed effects. The dependent variable is our standardized measure of nationalism. The full results are available in Table A3

less about immigrants, and have weaker priors about them. Local experience/contact shapes their preferences more significantly in either direction. Conversely, highly-educated people are likely to have been more-heavily exposed to different cultures and ethnicities (e.g. through travel, higher education, etc..), hence local contact with immigrants does not change much their views, which are, on average, positive toward immigration.

Age is a relevant dimension, too, in affecting citizens' responses to immigration (see panel

(b)). Size and statistical significance of the marginal effect of less-skilled migration on nationalist voting increase somewhat with age, though the response to high-skilled immigration does not seem to change monotonically with age and is stronger and more statistically significant for the age groups 18-37 and 38-57. This probably reflects a higher frequency of migrant-native interactions in these age categories (e.g. for work or study reasons). Panel (c) displays results of the heterogeneity exercise by gender. Immigration (both less- and high-skilled) affects nationalist voting mostly among males. The voting behavior of females is much less sensitive to immigration, being more similar to the voting behavior of highly-educated individuals. Moreover, females are also less likely than males to vote for nationalistic agendas, on average. Finally, panel (d) reports results by voters' place of residence. The strongest response in either direction is concentrated among individuals living in mid-size towns compared to cities.[27] This may reflect the fact that an increase in the number of migrants in a city, where exposure to migration is already high, does not add much to the information set of local voters. Nationalist voting behavior of individuals living in towns is encouraged by less-skilled migration. This suggests that frequency of interactions or intensity in local news may play a bigger role in small towns. Voters in rural areas respond to less-skilled migration by voting for more nationalist parties, while they do not seem to respond to high-skilled migration.[28]

3.5.2 Estimates of Differential Regional Effects

In Table 3.7 we show the interaction between broad regional characteristics and the impact of immigrants on nationalist voting. We split regions according to three indicators of economic and social outcomes: the value of per-capita social benefits (not in-kind) as a measure of welfare generosity; the ratio of children to adults at the regional level, as a proxy of the relevance of school as a local public good;[29] and the total number of crimes per person in the region as a proxy for public safety.[30] For each variable, we split the sample of regions above and below the median value in 2010. We then estimate the skill-specific impact of immigrants on individual voters' preferences on the two sub-samples of regions above and

[27]Similar results were found by Barone *et al.* (2016), where the effect of immigration on voting shares for right-wing parties in Italy are mainly driven by middle sized municipalities.

[28]Table A3 in the appendix reports the full set of results.

[29]Halla *et al.* (2017) show that the presence of immigrants could affect the quality of local schools and that could influence voting preferences of parents who are worried about the education of their children. Also, in Tables A15 and A16, we carry out a similar heterogeneity exercise for regional unemployment conditions.

[30]The proxy of crime level is computed as the total number of crimes, including robberies, homicides, burglaries and thefts, per 100,000 people. See the online 3.9.2 for details on source and construction of these variables.

below the median value of each regional characteristic, respectively. If the broad regional context is relevant to explain the effect of immigration on individual voting behavior, our empirical approach should be able to capture it.

Table 3.7: Nationalism and Immigrant Share - Regional Subsamples

	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	IV	IV	IV	IV
Time	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016
Regional Charact.	Social Transfer		Children		Crime	
Median	Below	Above	Below	Above	Below	Above
Share HS	-0.25***	-0.08	0.08	-0.25***	-0.14	-0.26***
	(0.08)	(0.09)	(0.08)	(0.06)	(0.22)	(0.07)
Share LS	0.10*	0.12	0.02	-0.07	0.31	0.09**
	(0.05)	(0.09)	(0.03)	(0.09)	(0.39)	(0.04)
Observations	23633	24670	24871	23432	17438	30865
Regions	55	59	58	56	55	59
K-P rk Wald F-stat	10.53	3.43	1.13	3.01	0.55	13.80
Adj. R-Square	0.16	0.09	0.13	0.13	0.13	0.13
NUTS2 f.e.	✓	✓	✓	✓	✓	✓
Year f.e.	✓	✓	✓	✓	✓	✓
NUTS2 Controls	✓	✓	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓	✓	✓

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Standard errors are clustered at NUTS2 level. * p<0.1, ** p<0.05, *** p<0.01. As individual controls age, gender and education dummy are included. As NUTS2 controls we include log GDP per capita, unemployment rate and share of tertiary educated individuals. The dependent variable is our standardized measure of nationalism. Analysis by subsamples at regional level based on social trasfer per capita (col (1) and (2)), on the ratio $\frac{0-14}{15-65}$ old (col (3) and (4)) and on the total number of crime (col (5) and (6)).

In columns (1) and (2), we show the IV estimates in regions with social transfers below/above the sample median. In columns (3) and (4) we show those with children-to-adult ratios below/above the sample median. Finally, in columns (5) and (6) we show the estimates in regions with crime rates below/above the sample median. The first finding is that, in general, the point estimates of high-skilled immigrant-share on nationalism is negative and the effect

of less-skilled immigrant-share is positive for most sub-groups.[31] The only significant effects of less-skilled immigrant-share is estimated in regions with high-crime levels, suggesting a stronger concern about less-skilled migrants in regions with high level of criminality. For high-skilled immigrants, their negative effect on nationalism is stronger in regions where social transfers are low and where crime is high. Possibly, the perception of a positive contribution from high-skilled immigrants is stronger in regions starting from low social transfers, crowded schools and high crime levels. One very important caveat, however, is that, when splitting the sample, the IV can be very weak (see the F-stat) and subject to bias. We should not attach excessive confidence to some of these estimates.

3.6 Effects on Turnout and on Longitudinal Individual Preferences

3.6.1 Decomposing the Effect: Turnout and Voting Preferences

The reduced form structure of equation (3.2), estimated on a repeated cross section, does not identify the mechanisms connecting migration with natives' voting behavior. One way to think about the mechanisms of such effects is to consider that local interactions with immigrants may produce a negative/positive revision of the degree of individual nationalism. In addition, the local presence of immigrants may contribute to make "immigration" a more salient issue in the voting decisions of some groups: people may vote for a party because of its position on immigration. In this sense, immigrants' presence may affect the nationalistic view of a person and/or make it more likely that those views are central in their vote. This channel has been emphasized by studies that show that attitudes are not easy to change, but events, information and news may change the willingness to act upon these attitudes with one's vote (see Barone *et al.*, 2015).

We very simply formalize this insight by considering two groups a and b, with different average preferences for the intensity of "nationalism" of their elected officials. These groups can be thought of as the group of skilled and less-skilled citizens, or young and old, or urban and rural. At period $t = 0$, the two groups have their own level of preferences towards nationalism, and no particular assumption is made on the relation between the two (i.e. $n_a^0 \gtreqless n_b^0$).

[31] Results with origin-specific shares of immigrants are available in Table A14 in the online appendix.

Next, we define a revision of average preferences for each group as a function of the local change in high- and less-skilled immigrants as shares of the population. This variable proxies the change in likelihood/frequency of interactions of the local population with immigrants and/or the change in intensity of local news related to them. Hence, the preferences of natives in period $t = 1$ may change as a result of immigrants' flow to a region. This is captured by the following relation:

$$n_j^1 = n_j^0 + f_j\left(\frac{\Delta s_m}{Pop}, \frac{\Delta u_m}{Pop}\right), \quad \text{for} \quad j = a, b. \tag{3.5}$$

In equation (3.5), the function $f_j(.,.)$ captures the revision of nationalism in group j as a consequence of being exposed to a change in skilled (s_m) and less-skilled (u_m) immigrants.

The basis for such update can be varied. Contact theory predicts that people fear less what they know (see e.g. McLaren, 2003). Highly-skilled immigrants are likely to be experienced as contributing citizens. They are likely to work and unlikely to be involved in "negative news stories". In that case, the derivative of the function with respect to skilled immigrant change would be negative ($f'_{j,s_m} < 0$). Stereotypical thinking may induce natives to attribute issues of social cohesion or economic problems to the presence of less-skilled immigrants. These immigrants may be perceived to be less likely to work and to be net users of local welfare. They may also be more likely to be in negative news stories, even if it is only a very small percentage of them. In this case, a larger inflow of less-skilled may produce a revision in favor of nationalistic views ($f'_{j,u_m} > 0$). A recent paper by Alesina *et al.* (2018) shows that people in many countries overestimate the share of immigrants and the share of non-educated ones in the population. They also show that for some groups of citizens, such as the less-educated ones in less-skilled occupations, this distortion in perception is even stronger.

The extent of the revisions may vary from group to group depending on several features. If the reason for the revision is a *Bayesian updating* of the contributions of immigrants, then groups who are initially less informed would have larger revisions. Existing literature (e.g. Della Vigna and Gentzkow, 2010) suggests that this can be the case for less-educated natives. These individuals might respond to local immigration more than do educated citizens. Alternatively, if the revision is driven by *confirmation bias* (so that signals confirming our beliefs have stronger effects), groups who are more nationalistic to begin with will respond to immigrant inflows–especially of less-skilled migrants–increasing their nationalism. These groups will use the new pieces of information to confirm their political priors (e.g. Nyhan and Reifler, 2010). Thus, they will not reduce their nationalism much in response to high-skilled immigration.

The other crucial factor in affecting electoral outcomes is the probability of actually voting. This can be affected by the presence of skilled or less-skilled immigrants, by influencing the "salience" of the immigration topic in deciding whether to vote. Defining the function $\delta_j\left(\frac{\Delta s_m}{Pop}, \frac{\Delta u_m}{Pop}\right)$ for $j = a, b$ as the probability of voting of individuals of group j, we can analyze the impact of immigration on it.

The index of political nationalism for the population is produced by aggregating the voting behavior and turnout of the two groups of individuals, weighted by their relative population. We standardize the index to one and denote by A and $1 - A$ the relative sizes of group a and b, respectively:

$$\begin{aligned}\overline{n^1} =& A\delta_a\left(\frac{\Delta s_m}{Pop}, \frac{\Delta u_m}{Pop}\right)\left[n_a^0 + f_a\left(\frac{\Delta s_m}{Pop}, \frac{\Delta u_m}{Pop}\right)\right] + \\ & (1-A)\delta_b\left(\frac{\Delta s_m}{Pop}, \frac{\Delta u_m}{Pop}\right)\left[n_b^0 + f_b\left(\frac{\Delta s_m}{Pop}, \frac{\Delta u_m}{Pop}\right)\right]. \end{aligned} \tag{3.6}$$

Immigration affects the index through the turnout channel and through the individual preference channel, and each channel can be specific to a group. Using our data we will be able to investigate these channels by estimating separately how the turnout of each individual group $j = a, b$ is affected by immigration, and how its preferences are affected.

In terms of the turnout channel we can estimate the following:

$$Turnout_{i,r,e} = \alpha + \beta_H m_{r,e}^H + \beta_L m_{r,e}^L + \zeta \mathbf{X}_{i,r,e} + \gamma \mathbf{\Gamma}_{r,e} + \theta_e + \eta_r + \epsilon_{i,r,e}. \tag{3.7}$$

The main regressors in equation (3.7) are analogous to equation (3.2) above. These are the shares of skilled (s) and less-skilled (u) migrants relative to the population in 2005 in region r. The outcome of interest is the probability that individual i, resident in region r votes in national election e. The explanatory variable of interest is the change of $m_{r,e}^s$ instrumented by shift-share-driven changes in $\frac{\Delta s_m}{Pop}$ and $\frac{\Delta u_m}{Pop}$. The coefficients of interest are $\hat{\beta}_s$ for $s = H, L$.

The second channel shown in equation (3.6) is the change in individual preferences. The information available from the ESS provides insight over this channel, too, by looking at longitudinal changes in individual political preferences between the last national elections and the current interview date.[32] Defining $Nation_{i,r,t}^{\pi^c}$ as the normalized level of nationalism

[32]The question in the ESS is defined as follows: “Which party do you feel closer to?". As it is shown in Table A7 in the online appendix, the percentage of "don’t know" answers is significant (on average 34%).

of party π^c preferred by voter i in region r in year t and $Nation^{\pi}_{i,r,t-1}$ as the normalized level of nationalism of party π voted-for in election year $t-1$, we can then compute:

$$\Delta Nation_{i,r,t} = Nation^{\pi^c}_{i,r,t} - Nation^{\pi}_{i,r,t-1}, \tag{3.8}$$

This is the variation in nationalism for individual i capturing a change in expressed preferences, conditional on having voted at $t-1$. Using expression (3.8) as our dependent variable, we can estimate an individual difference analysis as follows:

$$\Delta Nation_{i,r,t} = \alpha + \beta_H \Delta m^H_{r,t} + \beta_L \Delta m^L_{r,t} + \theta_t + \Delta\epsilon_{i,r,t}, \tag{3.9}$$

where $\Delta m^s_{r,t}$ is the change in the (skilled or less-skilled) migrant share of the population in region r between the year t and the year of the last election $t-1$. This corresponds to a linearization of function f_j in equation (3.6). Let us emphasize that the individual panel analysis fully controls for individual characteristics and time-invariant regional factors. It also abstracts from issues of selection into vote participation (which is analyzed in the "turnout" equation) as it includes all people who voted in the early election independently of their later participation. Poutvaara and Steinhardt (2018) is the only paper that we know of which exploits variation in individual preferences over time. However, that study looks at the effect of "bitterness" on attitudes towards migration in Germany, while we analyze the effect of local migration on individual voting across Europe.

3.6.2 Decomposing the Effects: Estimates

Table 3.8 presents the estimated coefficients on the share of high- and less-skilled immigrants on voter turnout, defined as the probability of voting among individuals aged 18 or higher. The specification is similar to the one for nationalism, except that a dummy "voting at last election" is used as the outcome instead of the nationalism index.

In aggregate, neither less-skilled nor high-skilled migration seem to be correlated with the propensity to vote in elections. However the average effect could hide heterogeneous reactions of specific groups of voters. When we consider separate sub-samples by education (tertiary and non-tertiary) in columns (2) and (3) and age (18-37, 38-57, over 58+) in columns (4) to (6), we find that high-skilled immigration is associated with a significant increase in turnout among young voters (age group 18-37). Taken at face value, the estimated coefficient suggests that a one standard deviation increase in high-skilled immigration increases the probability of voting among youths by 15%. This result shows that the negative impact of high-skilled

Table 3.8: Voting Turnout and Immigrant Share

	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	IV	IV	IV	IV
Time	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016
Natives		*Education*		*Age Groups*		
Subsample	All	Not Tertiary	Tertiary	18-37	38-57	58+
Share HS	0.01	0.01	0.02	0.05**	-0.00	-0.01
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
Share LS	0.00	0.00	0.01	0.01	-0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	78814	53675	25139	20730	28605	29479
K-P rk Wald F-stat	20.35	13.99	23.21	22.85	14.21	19.25
Adj. R-Square	0.10	0.09	0.07	0.13	0.06	0.03

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Standard errors are clustered at NUTS2 level. * p<0.1, ** p<0.05, *** p<0.01. As individual controls age, gender, education dummy are included. As NUTS2 controls we include log GDP per capita, unemployment rate and share of tertiary educated individuals. The dependent variable is a dummy that take values of 1 if the individual voted during the last national elections, zero otherwise. From column (2) to column (6) we perform our analysis on subsample by education (col. (2) and (3)) and age groups (col. (4),(5) and (6)).

migration on nationalist voting is partly driven by an increase in participation among young voters, whose preferences tend to be less nationalist than average.[33]

We then move to investigate the effects of migration on the change in individual preferences, by exploiting a longitudinal component of the data. Focusing on the IV estimates shown in Table 3.9 we see a positive and significant effect of the change in the less-skilled migrant share on individual preferences for nationalism. This effect is more precisely estimated for less-educated voters (column (4)). High-skilled immigration does not affect individual preferences significantly. In columns (7) to (10) we split the sample of voters between "radical" and "not radical", defining as "radical" those who voted for a party with an high level of nationalism (which are, accordingly to Figure 3.3, Radical-Right, No Family, Agrarian/Center and Confessional parties). The estimates suggest that the pro-nationalism effect of less-skilled migrants is much larger among "radical" individuals. This suggests that less-skilled migration tends to confirm the political views of already nationalist voters. Column (10)

[33] Additional results are available in the online 3.9.8, where we explore the effects by political affiliation of voters (Table A18). On average, we do not find any significant results.

shows a positive coefficient for high-skilled immigrants on non-radical voters, but its size is quite small.

Table 3.9: Nationalism and Immigrant Share - Individual Panel Analysis

	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV	(9) OLS	(10) IV
Time	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016
Natives	*All*		*Not Tertiary*		*Tertiary*		*Radical*		*Not Radical*	
Share HS (Δ)	0.03**	0.02	0.02*	0.01	0.04**	0.04	0.01	0.05	0.01	0.04*
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.03)	(0.20)	(0.60)	(0.01)	(0.02)
Share LS (Δ)	0.02**	0.05***	0.02*	0.06***	0.02**	0.03	0.38***	0.87*	0.01	0.03*
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.03)	(0.10)	(0.45)	(0.01)	(0.02)
Individuals	29238	29238	18229	18229	11009	11009	2320	2320	24244	24244
K-P rk Wald F-stat		6.05		5.58		6.13		1.41		6.44
Year f.e.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Standard errors are clustered at NUTS2 level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is the difference between our standardized measures of nationalism of the party that they feel close to and the party for which they voted for in the last national election. The analysis is performed over the whole immigrants shares population. The analysis is performed on different samples of natives voters based on their level of education and party voted for during the last election: all individuals (col (1),(2)), not tertiary educated (col (3),(4)) tertiary educated (col (5),(6)), voted for a "radical" party (col (7),(8)) and voted for a "not radical" party (col (9),(10)). The estimation is performed trough OLS and IV estimation procedure. 'Radical" parties are defined by their political family: radical right, no family, agrarian/center and confessional (see Figure 3.3). We drop Switzerland from the sample from columns (7) to (10), since we do not have any information of the political families of its parties.

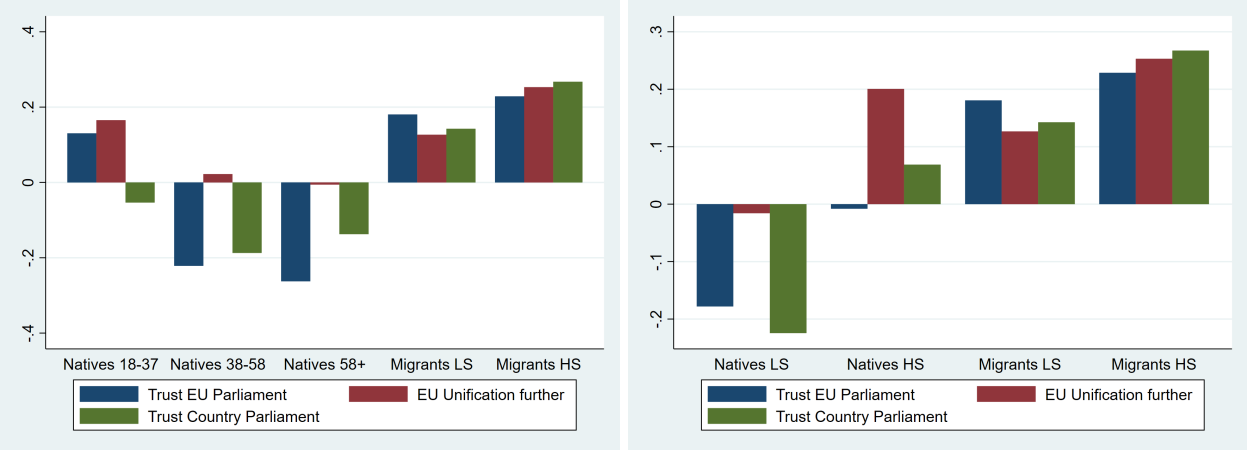
These results on individual responses to immigration suggest that a large part of the effect of less-skilled immigrants on voters in the repeated cross section comes from changes in individual preferences. The point estimates for individuals are not far from the aggregate ones shown in Table 3.4.[34] Caution is needed when interpreting the individual results. In general, a statement of current party preference in a survey may be an imprecise indicator of party voting. More importantly, individuals are better informed about party platforms when national elections are close than at any random time between them. This may generate a larger measurement error of preferences in the current period, which could result in an attenuation of the effects. The subsample we use for estimates in Table 3.9 is also smaller than our baseline sample, as only the 60% of individuals who voted in national elections also responded to the question about party closeness.[35] With these caveats in mind, we can say that results in Tables 3.8 and 3.9 are consistent with the following explanation of aggregate results. First the negative effect of high-skilled migration on nationalist voting

[34]The coefficient for less-skilled migrant-share in column (2) indicates that a one standard deviation increase in less-skilled migration (0.97) increases the variation of individual nationalism by $0.049 (= 0.97 * 0.05)$, that is, by roughly 8% of a standard deviation of the variation of nationalism indicator (0.628).

[35]In Figure A2 in the online appendix we show that the nationalism distribution of individuals in the subsample responding to this question is the same as in the main sample. Moreover, estimating equation (3.2) using the subsample of people that express the party voted-for during the last national election and the party they actually feel close to, we obtain the same results as in Table 3.4. This reassures us that the results in Table 3.9 are not driven by selective attrition into non-response in the current period.

may be mainly associated with increased political participation of young voters. Second, the positive effect of less-skilled migration on nationalist voting is mainly due to individual preference shifts, which mainly affect less-informed and less-educated voters and reinforces radical political views.

Figure 3.11: Political Attitudes of Natives and Immigrants



(a) Age groups (b) Education groups

Note: authors' calculations on ESS. The figures plots the average position by groups of the population, splitting them by natives age and education groups and migrants education groups. Figures (a) and (b) show the average trust on the EU parliament, trust on the country parliament and the will of a stronger EU unification by age and education groups. Each variable is standardized with mean zero and standard deviation equal to one.

Can part of the anti-nationalist impact of high-skilled immigrants be due to peer effects on political attitudes? To dig somewhat into this channel, in Figure 3.11, we show the attitudes toward political institutions of natives and immigrants. The figure represents indicators of trust towards the European parliament, the national parliament, and willingness to pursue EU unification for natives, by age group (panel (a)) and skill level (panel (b)), compared with the corresponding attitudes of less- and high-skilled immigrants. The Figure suggests that immigrants, and especially highly-educated ones, are characterized by the highest level of trust toward the EU and local political institutions. This level is higher than for all group of natives, including highly-educated and young ones.

We next analyze the relationship between immigration and natives' trust of institutions. In Table 3.10 we look at the effect of less- and high-skilled immigrants on natives' level of trust towards the EU parliament (columns (1) and (2)), natives' preferences for further EU unification (columns (3) and (4)), and trust towards their country's parliament (columns (5) and (6)). We report OLS and IV estimates using the usual shift-share approach. The results

Table 3.10: Immigration and Attitudes towards Politics

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	OLS	IV	OLS	IV
Time	2010-2016	2010-2016	2010-2016	2010-2016	2010-2016	2010-2016
Dep. Var.	*Trust EU Parl.*		*More EU*		*Trust Country Parl.*	
Share HS	0.03***	0.06*	0.02	-0.03	0.02**	0.07**
	(0.01)	(0.03)	(0.02)	(0.06)	(0.01)	(0.04)
Share LS	0.01	0.00	-0.00	0.00	0.01	-0.01
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.03)
Observations	73935	73935	56363	56363	78058	78058
K-P rk Wald F-stat		4.10		1.76		4.31
Adj. R-Square	0.08	0.08	0.10	0.10	0.10	0.10
NUTS2 f.e.	✓	✓	✓	✓	✓	✓
Year f.e.	✓	✓	✓	✓	✓	✓
NUTS2 Controls	✓	✓	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓	✓	✓

Note: authors' calculations on ESS, EULFS and Eurostat data. Standard errors are clustered at NUTS2 level. * p<0.1, ** p<0.05, *** p<0.01. As individual controls age, gender and education dummies are included. As NUTS2 controls we include log GDP per capita, unemployment rate and share of tertiary educated individuals. The dependent variables are different standardized measure of trust and attitudes towards politics: trust in the EU parliament (col. (1),(2)), EU unification should go further (col. (3),(4)) and trust in the national parliament (col. (5),(6)).

show that high-skilled migration increases native trust towards the parliamentary process, both at the EU and the national level, while less-skilled immigrants do not significantly affect them. These results are consistent with the idea that high-skilled immigrants may impact native voting in part by making natives more similar to the high-skilled immigrants themselves.

3.7 Simulations of Nationalist Voting Scenarios

Our results provide a new understanding of how local inflows of less- and high-skilled immigrants interact with characteristics of local residents and affect their intensity of nationalist

voting behavior. In particular the education levels of European citizens and of immigrants turns out to be very important factors in determining the voting response to immigration. This implies that both immigration and education policies will affect the politics of regions vis-a-vis immigration (see e.g. Braga *et al.*, 2013). In this section, we produce some simulations based on our estimates. They present different scenarios on the response of nationalism levels across European regions under alternative hypotheses regarding potential inflows of immigrants and the potential education level of the local population.

As a benchmark, we consider the regional change in levels of nationalism (of voted parties) predicted by the coefficients on the skill-specific migration shares estimated separately by education category (as in Figure 3.10, panel (a)). We first feed into this prediction the observed net changes in less-skilled and high-skilled migrants during the last ten sample years (i.e. over the period 2007-2016). In practice, we predict the regional response of nationalism to immigration in 2007-2016 using the following equation:

$$\Delta\overline{Nation_r} = \hat{\beta}_{NT}^{HS}(1 - Tert_{r,16})\frac{\Delta M_{r,16-07}^{HS}}{Pop_{r,05}} + \hat{\beta}_{T}^{HS}(Tert_{r,16})\frac{\Delta M_{r,16-07}^{HS}}{Pop_{r,05}} + \hat{\beta}_{NT}^{LS}(1 - Tert_{r,16})\frac{\Delta M_{r,16-07}^{LS}}{Pop_{r,05}} + \hat{\beta}_{T}^{LS}(Tert_{r,16})\frac{\Delta M_{r,16-07}^{LS}}{Pop_{r,05}}. \quad (3.10)$$

In equation (3.10), $\hat{\beta}_E^s$ is the estimated coefficient of the effect of a share of migrants with skill $s = HS, LS$ (high-skilled, less-skilled) on nationalism of natives with education $E = T, NT$ (tertiary, non-tertiary). Each coefficient is weighted by the corresponding share of natives with and without tertiary education ($Tert_{r,16}$) and multiplied by the observed change in the stock of migrants in the corresponding skill group, (normalized by population size in 2005 i.e. $\frac{\Delta M_{r,16-07}^{s}}{Pop_{r,05}}$)[36].

In Table 3.11, we report results from this simulation at the country level. Results in column (3) show that, on average, the combination of natives' education and skill composition of migration flows in the last ten years reduced nationalism levels in the EU. However there is a lot of cross-country heterogeneity. Column (2) shows country-specific average predicted effects, using the regional population as weights. The model predicts an increase in nationalism in countries characterized by relatively low shares of people with tertiary education

[36]Even though we could not include Italy and Greece in our estimations due to a lack of voting data, we include them in our simulations. By so doing, we assume that the response of their citizens is similar to that in other EU countries, and we can predict their simulated response as they are important immigration-receiving countries in the EU context.

Table 3.11: Simulated Effect of Immigration/Education Changes on Nationalism: Country-level and EU averages

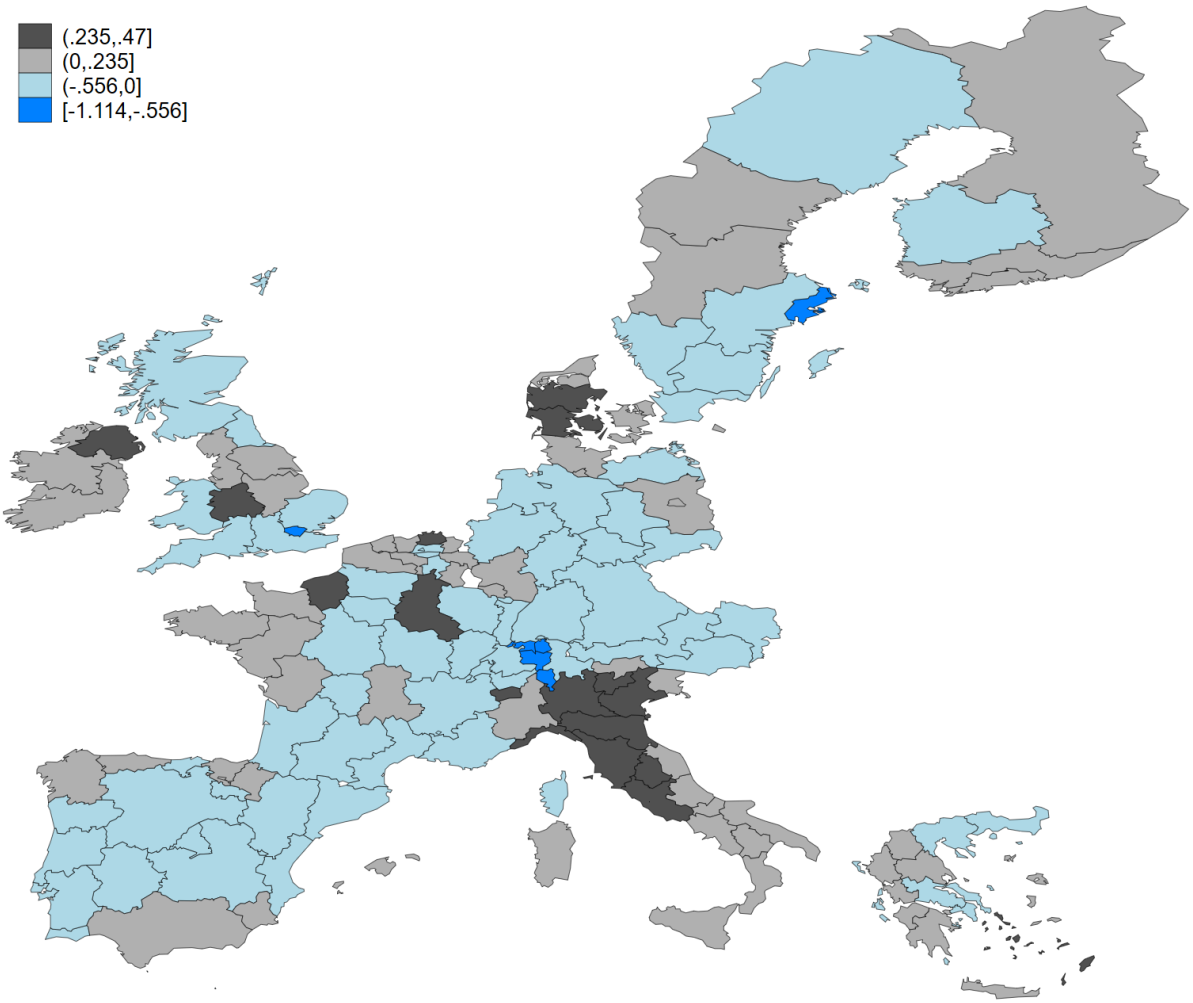
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Standard Sim.		Country Characteristics			Mig. Policies		Educ. Policies	
Country	$\Delta Nation_c$	$\Delta\widehat{Nation}_c$	$\Delta\widehat{Nation}_{EU}$	$\overline{Tert_c^{16}}$	$\overline{\Delta Mig_c^{HS}}$	$\overline{\Delta Mig_c^{LS}}$	*No LS*	*Balanced*	+20% *Tert.*	*Max Tert.*
Austria	0.04	-0.299	-0.032	0.314	2.641	1.597	-0.442	-0.16	-0.36	-0.364
Belgium	-0.018	0.042	-0.032	0.375	2.131	4.183	-0.304	-0.221	-0.053	-0.079
Denmark	-0.092	0.171	-0.032	0.377	0.47	2.898	-0.1	-0.143	0.068	0.052
Finland	0.222	0.002	-0.032	0.428	1.254	2.185	-0.119	-0.083	-0.1	-0.099
France	0.255	-0.102	-0.032	0.349	0.676	0.445	-0.166	-0.066	-0.177	-0.19
Germany	0.141	-0.051	-0.032	0.282	1.086	1.499	-0.209	-0.112	-0.117	-0.128
Greece	.	-0.046	-0.032	0.3	0.081	0.399	-0.021	0.007	-0.111	-0.122
Ireland	0.28	0.085	-0.032	0.43	0.477	2.081	-0.079	-0.091	-0.024	-0.005
Italy	.	0.259	-0.032	0.177	0.473	3.357	-0.086	-0.154	0.209	0.204
Portugal	0.043	-0.116	-0.032	0.239	0.728	0.107	-0.105	-0.019	-0.167	-0.175
Spain	0.045	-0.036	-0.032	0.357	0.339	0.178	-0.041	-0.011	-0.116	-0.13
Sweden	0.103	-0.268	-0.032	0.409	3.018	3.319	-0.517	-0.231	-0.349	-0.362
Switzerland	0.494	-0.508	-0.032	0.411	5.621	3.347	-0.794	-0.312	-0.574	-0.589
United Kingdom	0.55	-0.1	-0.032	0.421	1.779	2.739	-0.315	-0.171	-0.192	-0.223

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Column (1) the country averages of the regional differences in nationalism between the last and the first election available. Columns (2) and (3) show country averages and EU average, respectively, of the simulated standardized nationalism from the standard simulation using as regional weights the total population. Column (4) shows the country regional average of tertiary individuals. Columns (5) and (6) show the country average variation of migration share of HS and LS immigrants respectively over the 2007-2016 period. Column (7) shows the result of the simulation when we remove the variation of low educated immigrants. Column (8) shows the result of the simulation when we assume skill-balanced immigration. Column (9) shows the result of the simulation when we increase by 20% the share of tertiary educated natives by region. Column (10) shows the result of the simulation when each region has the same share of tertiary educated natives as the highest educated one in the same country.

(e.g. Italy) and/or a prevalence of less-skilled migration (as in Belgium, Denmark, Italy). On the contrary, predicted nationalism decreases in countries such as Austria, Sweden and Switzerland, as these are characterized by high shares of tertiary-educated individuals (more than 30%) and more balanced inflows of less- and high-skilled immigrants in the considered period. The share of tertiary-educated people in the country is reported in column (4), while the 2007-2016 inflow of high-skilled and less-skilled migrants as a percent of the group is reported in columns (5) and (6) of Table 3.11. Notice that immigration-predicted changes in nationalism in each country have quite different magnitudes and sometimes opposite signs compared to the actual changes presented in column (1) of Table 3.11. This makes clear that several other factors have affected the rise of nationalism, besides the actual inflow of immigrants (e.g. economic conditions, increased inequality, globalization, frustration with EU bureaucracy and so on. See Algan *et al.*, 2017 and Rodrik, 2018).

Figure 3.12 displays the predicted changes of nationalism at the NUTS2 regional level for all countries of Western and Northern Europe. In several regions, the skill composition and the migration inflows during 2007-2016 predict a decrease in nationalism (areas in light and dark blue). This is the case in regions such as Madrid (-0.123) in Spain, or Thuringia (-0.075) in Germany. These places are characterized by low levels of less-skilled immigration and

Figure 3.12: Baseline Simulation Analysis



Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. The figure plots the results of equation (3.10) at NUTS2 level, which simulate the average level of nationalism over 2007-2016 due to the variation of immigrants by education and the level of natives education. For Germany, United Kingdom and Austria the results are plotted at NUTS1 level. Azores islands, Madeira islands and Canary islands are not plotted for visual reasons.

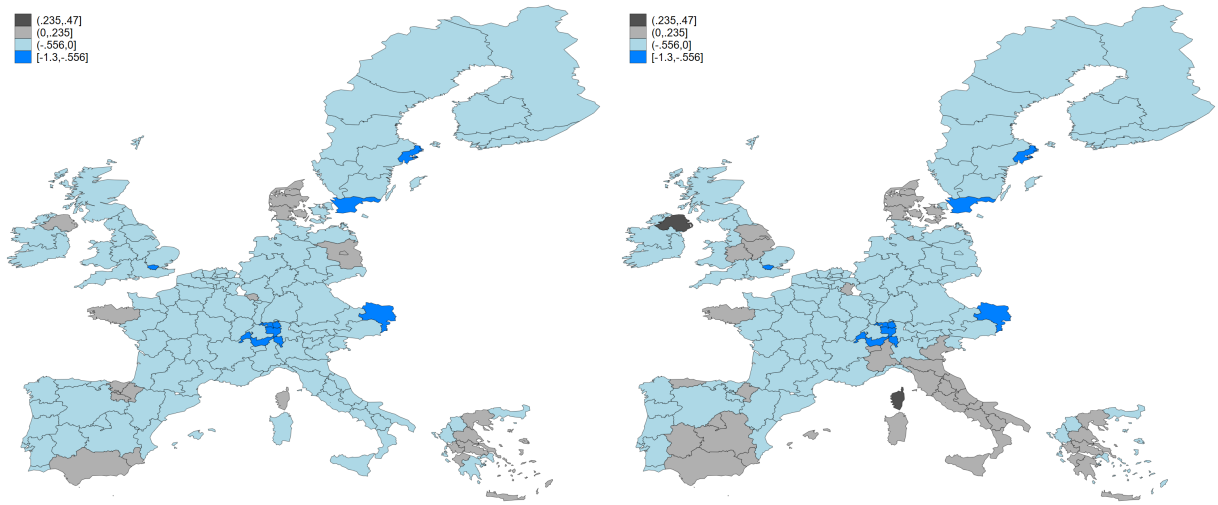
large net flows of highly-educated immigrants, combined with a significant share of tertiary-educated natives. Nationalism decreases even more in regions such as Provence-Alpes-Côte d'Azur (-0.399) in France, Zurich (-1.113) in Switzerland, and Greater London (-0.789) in the UK, where the share of high-skilled migrants was very large relative to less-skilled migrants[37]. This may have produced positive and trust enhancing migrant-to-native interactions. On the opposite side of the spectrum, we find areas that attracted many less-skilled migrants compared to high-skilled migrants. The rise in less-skilled immigration compared to high-skilled immigration may have worsened attitudes towards immigrants, and increased the level of nationalism in regions such as Basilicata (0.234) in Italy, West Midlands (0.329) in the UK, and Syddanmark (0.285) in Denmark. In these regions the rise of less-skilled immigration was more than 10 times higher than the rise of high-skilled immigration.

Figure 3.13 shows the results of our estimated model when we simulate four alternative scenarios of nationalist responses to different migration outcomes. The first scenario hypothesizes a *skill-selective migration*, obtained by setting $\Delta M^{LS} = 0$ (no less-skilled) in equation (3.10) while leaving ΔM^{HS} as on 2007-2016. Country-specific results for this scenario are available in column (7) of Table 3.11 and a map representation is in depicted in panel (a). Compared to the benchmark in Figure 3.12, this selective policy induces a decline of nationalism almost everywhere in Europe. Such a fall is significant and shows a different effect relative to the benchmark in regions of Italy, the north of France, and the UK. This scenario is discriminatory, and explicitly violates Schengen rules of free circulation of European citizens. In panel (b) we present a second scenario: a skill-selective policy that allows less-skilled migration from EU countries only, but prevents less-skilled non-Europeans from entering. This is obtained by setting $\Delta M^{LS}_{nonEU} = 0$ for all regions in equation (3.10). This policy would be less effective in preventing nationalism in regions in Italy, Spain and the UK, as it would not prevent less-skilled migration from Central European EU countries. The two scenarios above impose full restrictions to less-skilled migration and their effects are not surprising: they reduce the nationalistic shift, since less-skilled immigration enhances nationalistic positions. The next two scenarios allow less-skilled migration but they balance it with high-skilled migration. These are in line with more open policies towards immigrants. In panel (c) we consider a *skill-balanced migration policy* obtained by setting $\Delta M^{LS} = \Delta M^{HS}$ in equation (3.10) so that immigration would not alter the skill ratio of a region[38]. Interestingly enough, this policy induces a decline in nationalist voting which is quite sizable, although not as strong as the first policy (panel (a)). The effects of this

[37]In some cases the variation of less-skilled immigrants is negative, like in Provence-Alpes-Côte d'Azur.

[38]For each region we do not change the size of the immigrant population, but we split between high- and less-skilled migrants so as to mirror native ratios.

Figure 3.13: Simulations for Different Migration Scenarios



(a) No LS Migration

(b) No Extra-European LS Migration

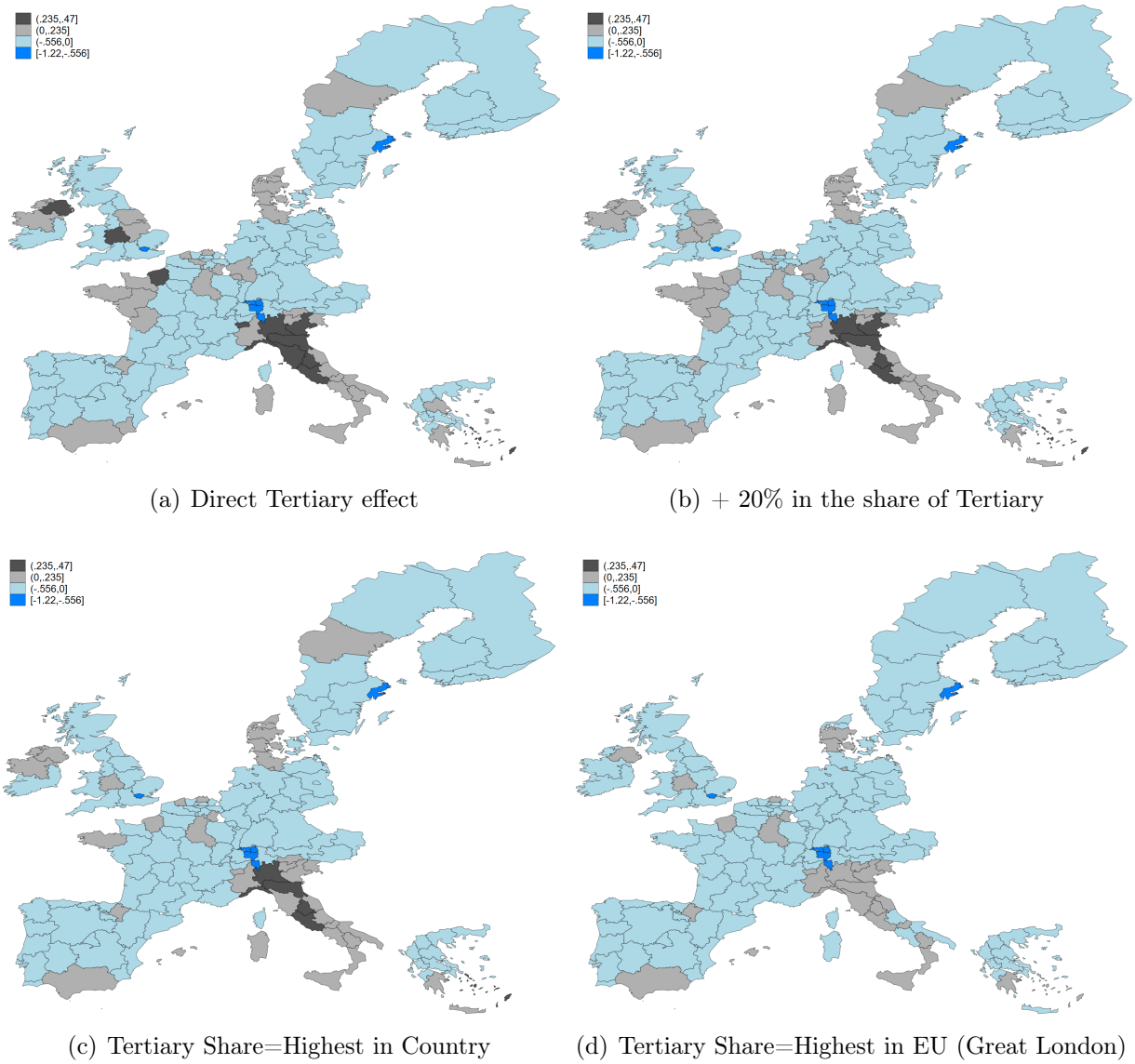
(c) Balanced Migration (LS=HS)

(d) Balanced Extra-European Migration (LS=HS)

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. The figure plots the simulated average level of nationalism after different migration policy scenarios, due to the variation of immigrants by education and level of education of natives over the 2007-2016 . The figure plots the results of equation (3.10) at NUTS2 level when: LS immigrants are excluded (panel (a)), not European LS immigrants are excluded (panel (b)), LS and HS immigrants are exactly balanced in each region (panel (c)) and not European LS and HS immigrants are exactly balanced in each region.

policy are also shown in column (8) of Table 3.11 for each country. In panel (d) we consider skill-balanced migration from non-EU countries only, i.e. $\Delta M^{LS}_{nonEU} = \Delta M^{HS}_{nonEU}$. Again, we see anti-nationalistic shifts in most regions by balancing the skilled- to less-skilled inflow from outside the EU.

Figure 3.14: Simulation for Different Education Scenarios



(a) Direct Tertiary effect

(b) + 20% in the share of Tertiary

(c) Tertiary Share=Highest in Country

(d) Tertiary Share=Highest in EU (Great London)

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. The figure plots the simulated average level of nationalism after different education policy scenarios, due to the variation of immigrants by education and level of education of natives over the 2007-2016. The figure plots the results of equation (3.10) at NUTS2 level when: the direct effect of tertiary education on nationalism is included (panel (a)), regional level of tertiary educated natives is increased by 20% (panel (b)), each region has the same level of tertiary educated natives as the highest region in the same country (panel (c)) and each region has the same level of tertiary educated natives as the highest in our sample (Greater London) (panel (d)).

Figure 3.14 presents, instead, four alternative predictions based on alternative education achievements in EU regions. In panel (a), we present a scenario that includes the direct effect of a higher share of tertiary-educated native individuals on nationalism. In practice, we add the term $\hat{\beta}_T Tert_{r,16}$ to the predictive equation (3.10). The coefficient $\hat{\beta}_T$ measures the direct effect of a higher share of tertiary educated citizens in the region on nationalism, as we estimated in individual regressions. This simulation scenario is not too different from the benchmark in Figure 3.12. It shows somewhat lower nationalist voting in northern regions of Spain, Continental and Northern European regions in Germany, Sweden, and Finland as the generalized increase in college-educated natives produced such direct attenuation. Education is an effective means to contrast nationalist attitudes in most of these regions. In panel (b), we then hypothesize a European education policy that raises the share of tertiary-educated individuals by 20% in all regions, and we keep immigration flows as they were in 2007-2016. Results are available in column (9) of Table 3.11. This policy reduces voting for nationalistic parties in the northern regions of Spain and the Continental and Northern European regions. It is much less effective in the UK and Southern European regions (e.g. in Italy, Spain and Greece), where nationalistic views are mostly driven by the presence of less-skilled migration. In panel (c) we present the effects of an alternative education policy that increases in each region the share of tertiary-educated individuals up to the highest regional share in the country. Results from this policy are qualitatively similar to the previous one. There are fewer votes for nationalist parties in the usual set of Continental and Northern European regions and Northern Spanish regions. Relative to simulations in panel (b), nationalism also decreases in Northern Greek regions. The scenario does not reduce nationalism levels in the UK, Italy and Southern Spain. Finally, in panel (d) we simulate the effects of a similar education policy that increases everywhere the share of tertiary-educated individuals up to 57%, which is the highest regional share in our sample (that we observe in Greater London). Results are available in column (10) of Table 3.11. Compared to simulation in panel (c), this policy moderates nationalism voting in some Italian regions and produces a significant reduction of nationalism in response to immigration in almost any place in Europe. However, nationalism still increases in this scenario in some Italian regions such as Lombardy, Latium, Umbria and Emilia Romagna.[39] Overall, balanced immigration seems a good antidote to

[39]Country-specific results not available in Table 3.11 are then available in Table A20 in the online appendix. Moreover, in the online 3.9.11 we perform other simulations and robustness checks. First, using only the statistically significant coefficients from our estimations, (i.e. keeping in equation (3.10) only non-tertiary-educated natives) we plot the results in Figure A3 panel (a). Second, in Figure A3 panel (b) we plot the results using origin-specific coefficients and origin-education-specific migrant variation. Results are similar to our baseline. Finally, in Table A21 and Figure A4 we show the level of nationalism after our predicted variation. Interestingly, several regions that experienced a strong variation towards nationalism were originally less nationalistic.

nationalistic voting. Policies promoting and encouraging high-skilled immigration in Europe would muster political support, to also allow for a certain amount of less-skilled migration.

3.8 Conclusions

In this paper we investigated the effects of different types of immigration on European voters' political preferences for nationalism. We first estimate an aggregate effect across regions, and then we analyze heterogeneous effects, potential channels, and the margins of participation and individual ideology in producing such effects.

We addressed these questions using a sample of elections in 12 European Countries over the period 2007-2016, and measuring the change in high- and less-skilled immigrants as shares of the adult population in regions within those countries. In order to directly address the connection between party-vote and nationalism, we classified each party on a common scale in terms of the intensity of their "nationalism", based on their political Manifesto. We linked this information to the current political preferences (in terms of party) and the party voted for in the last election, for a large sample of citizens in 12 EU countries. A skill-specific shift share based on the location of immigrants by country of origin in 2005 provides the IV to proxy push-driven changes in immigrant populations. The significant differentiation in country-of-origin and skill intensities of groups of immigrants allowed us to estimate two separate effects for skilled and less-skilled immigrants on residents' voting preferences.

We found a robust negative effect of high-skilled immigration on the intensity of nationalist preferences of citizens. We also found an opposite, positive association of approximately the same magnitude between less-skilled immigration and citizens' nationalist preferences. We also uncovered that the change in preferences driven by immigrants is stronger for less-educated citizen voters than for the highly-educated. Our evidence is consistent with the view that these opposite effects of less- and high-skilled migration on nationalism may be driven by different channels: the *pro-nationalist effect of less-skilled immigration* arises as natives, especially less-skilled ones, reinforce their negative priors regarding the less-skilled immigrants and become more pro-nationalism. The *anti-nationalist effect of high-skilled immigration* may be instead triggered by interactions and peer effects of high-skilled immigrants, who increase political trust of natives and push the more anti-nationalist young cohorts to vote in greater numbers.

We then presented a number of simulations regarding potential scenarios implying different migrant combinations and different levels of native educational attainment. These simula-

tions point to three main facts. First, immigration policies producing more balanced inflows of less- and high-skilled migrants would shift political preferences away from nationalist voting. These policies would not reduce immigration, but re-balance it. Second, closing borders to less-skilled migrants would also be associated, in most cases, with lower nationalist voting. However, these would be discriminatory measures, explicitly against the Schengen rules. Finally, policies increasing the share of tertiary-educated individuals may reduce nationalist voting in some countries, particularly Northern European countries. This is because college-educated people tend to cast fewer votes for nationalistic parties, and their voting behavior responds less to changes in immigration. We hope these ideas will stimulate further research in this area.

3.9 Appendix

3.9.1 Additional Tables

Table A1: Nationalism Intensity and Immigrant Share, Initial and Final Party Positions

	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
Time	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016
Panel A - Initial Position								
Share	-0.02	-0.00						
	(0.01)	(0.04)						
Share HS			-0.08***	-0.13**			-0.09***	-0.16***
			(0.03)	(0.06)			(0.03)	(0.06)
Share LS					-0.00	0.03	0.02	0.05*
					(0.01)	(0.04)	(0.01)	(0.03)
Observations	48219	48219	48219	48219	48219	48219	48219	48219
K-P rk Wald F-stat		45.24		31.38		38.54		14.67
Adj. R-Square	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
Panel B - Final Position								
Share	-0.01	0.00						
	(0.01)	(0.04)						
Share HS			-0.06**	-0.12*			-0.07***	-0.15**
			(0.03)	(0.06)			(0.03)	(0.06)
Share LS					0.01	0.04	0.02*	0.06**
					(0.01)	(0.04)	(0.01)	(0.03)
Observations	48225	48225	48225	48225	48225	48225	48225	48225
K-P rk Wald F-stat		45.25		31.54		38.53		14.74
Adj. R-Square	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
NUTS2 f.e.	✓	✓	✓	✓	✓	✓	✓	✓
Year f.e.	✓	✓	✓	✓	✓	✓	✓	✓
NUTS2 Controls	✓	✓	✓	✓	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Standard errors are clustered at NUTS2 level. . * p<0.1, ** p<0.05, *** p<0.01. As individual controls age, gender and education dummy are included. As NUTS2 controls we include log GDP per capita, unemployment rate and the share of tertiary educated individuals. The dependent variable is our standardized measure of nationalism measured in different moment: as the first observation before the beginning of the sample period (Panel A) and the last observation available by party (Panel B). These compare with the average over the sample period, which is used as a dependent variable in baseline estimates in Table 3.4

3.9.2 Descriptive statistics and data sources

We provide additional descriptive statistics as well as additional details on the data sources and definitions. Section 3.9.2 provides information of the different data sources. Table A5

Table A2: Salience of Nationalism and Immigrant Share

	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
Time	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016
All Migrants								
Share	0.000	-0.000						
	(0.000)	(0.001)						
Share HS			-0.001	-0.004***			-0.001*	-0.004***
			(0.001)	(0.001)			(0.001)	(0.001)
Share LS					0.000	0.001	0.001***	0.001**
					(0.000)	(0.001)	(0.000)	(0.001)
Observations	48319	48319	48319	48319	48319	48319	48319	48319
K-P rk Wald F-stat		45.42		32.24		38.71		14.96
Adj. R-Square	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
Not EU Migrants								
Share No EU	0.000	0.001						
	(0.000)	(0.001)						
Share No EU HS			-0.001	-0.007**			-0.002*	-0.011***
			(0.001)	(0.003)			(0.001)	(0.004)
Share No EU LS					0.000	0.001	0.000*	0.003***
					(0.000)	(0.001)	(0.000)	(0.001)
Observations	48319	48319	48319	48319	48319	48319	48319	48319
K-P rk Wald F-stat		40.48		39.42		44.70		17.00
Adj. R-Square	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.11
NUTS2 f.e.	✓	✓	✓	✓	✓	✓	✓	✓
Year f.e.	✓	✓	✓	✓	✓	✓	✓	✓
NUTS2 Controls	✓	✓	✓	✓	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Standard errors are clustered at NUTS2 level. * p<0.1, ** p<0.05, *** p<0.01. As individual controls age, gender and education dummy are included. As NUTS2 controls we include log of GDP per capita, unemployment rate and share of tertiary educated individuals. The dependent variable is our measure of salience of nationalism.

lists the variables used in the empirical analysis with their definition and source. Section 3.9.2 describes the voting data available from the ESS.

Data description

Our analysis is built on four different datasets: (i) European Social Survey (ESS), (ii) Manifesto Project Database (MPD), (iii) European Labor Force Survey (EULFS) and (iv) Eurostat. Here below we describe each data source, their characteristics and how we combine them.

Table A3: Nationalism and Individual Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV
Time	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016
Subsample	*Education*		*Age Group*			*Gender*		*Domicile*		
Natives	Not Tertiary	Tertiary	18-37	38-57	58+	Male	Female	Rural	Town	City
Share HS	-0.19**	-0.08	-0.22**	-0.21***	-0.12	-0.24***	-0.10	-0.07	-0.12	-0.10
	(0.09)	(0.07)	(0.11)	(0.07)	(0.09)	(0.09)	(0.07)	(0.09)	(0.12)	(0.07)
Share LS	0.11***	0.03	0.02	0.09**	0.10**	0.10**	0.05	0.12***	0.13**	0.01
	(0.04)	(0.04)	(0.05)	(0.03)	(0.05)	(0.05)	(0.03)	(0.05)	(0.06)	(0.03)
Observations	31104	17215	10849	17787	19683	24054	24265	18939	15088	14292
K-P rk Wald F-stat	11.04	21.04	13.75	13.13	15.21	14.55	14.83	4.42	8.88	13.62
Adj. R-Square	0.14	0.16	0.14	0.12	0.17	0.14	0.13	0.14	0.14	0.13
NUTS2 f.e.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year f.e.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
NUTS2 Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Standard errors are clustered at NUTS2 level. * p<0.1, ** p<0.05, *** p<0.01. As individual controls age, gender and education dummies are included. As NUTS2 controls we include log GDP per capita, unemployment rate and share of tertiary educated individuals. The dependent variable is our standardized measure of nationalism. Analysis by subsamples based on individual education groups (col. (1) and (2)), on age groups (col. (3)-(5)), on gender (col. (6) and (7)) and place of residence (col. (8)-(10)).

European Social Survey (ESS)- Our primary data source is the European Social Survey. This is a multi-country survey, which was administered in 8 waves (one every two years) in 36 countries between 2002 and 2016. Each individual in the ESS is selected by strict random probability method and the samples are representative of the population over 15 in each country. On average each wave contains around 1500 individuals for each country. The data include detailed information on personal and family characteristics such as age, gender, education, marital status, number of children in the family, place of birth and labor market characteristics such as employment status and work characteristics. It also includes detailed information on the parental background, such as parents' education, employment status, occupation when the respondent was 14 years old and their country of birth. From 2010 on it provides also intra-country geographical location of respondents at NUTS2 level. Finally, it provides also voting and political preferences of individuals. In particular, we are interested in two specific questions: (i) which party did you vote in the last national election? (ii) which party do you fell close to? The answers to those questions are the actual names of the parties in each surveyed country, giving a clear definition of the voting preferences/political closeness of each individual.

Manifesto Project Database (MPD) – The Manifesto Project Database, originally created by the Manifesto Research Group in the late 1970s and evolved under different names (e.g. Comparative Manifesto Project), analyses the parties' political manifesto to study parties'

Table A4: Attitudes towards Migrants and Immigrant Shares - All Native Respondents

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	OLS	IV	OLS	IV
Time	2010-2016	2010-2016	2010-2016	2010-2016	2010-2016	2010-2016
Attitudes	*Good Economy*		*Enrich Culture*		*Better place to live*	
Share HS	-0.01	0.01	0.00	0.00	-0.01	-0.03
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
Share LS	-0.00	-0.02	-0.01	-0.02	-0.01	-0.04**
	(0.01)	(0.03)	(0.01)	(0.02)	(0.01)	(0.02)
Observations	74072	74072	74085	74085	74091	74091
K-P rk Wald F-stat		3.96		3.96		3.96
Adj. R-Square	0.13	0.13	0.13	0.13	0.12	0.12
NUTS2 f.e.	✓	✓	✓	✓	✓	✓
Year f.e.	✓	✓	✓	✓	✓	✓
NUTS2 Controls	✓	✓	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓	✓	✓

Note: authors' calculations on ESS, EULFS and Eurostat data. Standard errors are clustered at NUTS2 level. * p<0.1, ** p<0.05, *** p<0.01. Individual controls are age, gender and education dummy are included. NUTS2 controls are log GDP per capita, unemployment rate and share of tertiary educated individuals. The dependent variable are different standardized (mean 0, sd 1) measures of attitudes towards immigrants: good for the economy (col (1)-(2)), enrich country cultural life (col (3)-(4)) and make the country a better place to live (col(5)-(6)). Higher values imply positive attitudes.

political preferences. It covers all the parties that are candidate at the national elections and gain at least one seat in the parliament. Democratic countries in the OECD and Eastern Europe are covered, having a sample of 56 countries over the 1945-2017 period. The number of parties analyzed by the MPD are 1093 over 715 parliamentary elections. Parties' political preferences are studied through a content analysis of the political manifesto: the share of quasi-sentences related to a topic are calculated as a fraction over the whole political manifesto. For each topic the MPD identifies two measures: one of favorable/positive mentions and the other of unfavorable/negative mentions. Several topics are analyzed, like the role of military, constitutionalism, decentralization, market regulation, etc. For our research we focus on parties' political preferences on: (i) European Community/Union (ii) National way of life. The former takes into account all the mentions on the EU like the desirabil-

ity/opposition of expanding the EU or increasing EU competences. The latter contains all the mentions related to nationalism, patriotism, pride of citizenship, etc. For each one of the two topics we then compute two measures. One is a measure of saliency, computed as the sum of the mentions (both positive and negative) related to the topic in analysis. The other is a measure of favorable political position of the party, computed as the difference between positive and negative mentions of the topic. We compute the average over time of those indicators for each party, dropping all years before the 1990.[40]. Finally, we harmonized and merged the MPD with the voting/political preferences of individual from ESS through the name of the party voted and the year of elections. In this way we know also the political preferences of the party voted by each individual in the ESS.

European Labor Force Survey (EULFS) – The European Labor Forces Survey is a large household survey conducted over the 28 members of the EU, the 3 member of the EFTA (Switzerland, Norway and Iceland) and two candidate countries. Data are available from 1983 on and it is representative of the population above 15. Information related to age, employment status and education are available in this survey. Moreover, from 2005, a disaggregate variable of country of birth is available across the majority of the countries. Fifteen birthplace regions are recognized by the EULFS: natives, EU15, New European member state from 2004, new European member state of 2007/2013, EFTA, Other Europe, North Africa, Other Africa, Near and Middle East, East Asia, South and South East Asia, North America, Central America and Caribbean, South America and Australia and Oceania. Thanks to the latter variable we can easily recognize the native and immigrant population at NUTS2 level in each country. Using the microdata of European Labor Force Survey and focusing on 14 European countries (Austria, Belgium, Denmark, Finland, France, Germany, Greece, Italy, Ireland, Portugal, Spain, Switzerland, Sweden and United Kingdom) we compute the share of immigrants over the total population in 2005 at NUTS2 level as follow:

$$m^{O}_{r,s,t} = \frac{M^{O}_{r,s,t}}{Pop_{r,2005}} \tag{3.11}$$

where $m^{O}_{r,s,t}$ is the share of migrants and $M^{O}_{r,s,t}$ is the total stock of migrants in NUTS2 region r born in the group of origin countries $O \in \{All, NonEU28\}$, with skills $s \in \{All, HS, LS\}$ at time t. For the 14 countries in analysis we built those origin-skill specific shares of migrants at NUTS2 level, excluding Austria, Germany and United Kingdom, where we built them at NUTS1 level. Since Germany does not provide information on the birthplace of the

[40]In this way we focus on a period in which the European integration proceess becomes an faster In most cases, information at the party level do not go back that far.

foreign born individuals in their country, we impute the country of birth with the available information on nationality and we distribute the population of foreign born naturalized people using the shares by nationality within the foreign born population. NUTS2 information are available for Denmark from 2007. Our sample includes 146 regions over the 2005-2016 period.

Eurostat – Being the statistical office of the EU, Eurostat can provide several socio-economics data of the EU members. For our analysis, we utilize Eurostat to build a vector of relevant NUTS2 control variables for our main analysis. In particular, we extract data on GDP per capita, population density, unemployment rate and percentage of tertiary educated individuals. Since a measure of GDP per capita is not available for Switzerland, we extract this information from the Regional Economy Dataset available from the OECD. Moreover, we used Eurostat to extract relevant data at NUTS2 level to perform our heterogeneity analysis at regional level, like the value of social benefits other than social transfer in kind per capita, the ratio 0-14 over 15-65 years old population and a measure of total number of crimes.

Table A5: Data Sources and definitions

Variable	Description	Definition	Source
Individual			
$Nationalism_{i,r,t}$	Measure of Nationalism	First component from a PCA on the shares of favorable mentions on EU and Nationalism of each partys' political manifesto. It measures the level of Nationalism of the party voted by individual i in region r at time t.	Author's Calculation on the ESS and MPD data.
$Salience_{i,r,t}$	Measure of Salience of Nationalism issue	First component from a PCA on the shares of total mentions on EU and Nationalism of each partys' political manifesto. It measures the saliency of those topics of the party voted by individual i in region r at time t.	Author's Calculation on the ESS and MPD data.
$Age_{i,r,t}$	Respondent's age	Age of individual i.	ESS data.

$Tertiary_{i,r,t}$	Tertiary dummy	Dummy variable that takes value of 1 if individual i is tertiary educated	ESS data.
$Female_{i,r,t}$	Woman dummy	Dummy variable that takes value of 1 if individual i is a woman.	ESS data.
$Tertiary\ fath_{i,r,t}$	Father educational background dummy	Dummy variable that takes value of 1 if individual i's father is tertiary educated.	ESS data.

Regional

$m^{O}_{r,s,t}$	Share of migrants	Share of migrants in region r at time t of skill s from origin countries O over the 2005 population of region r.	Authors' Calculation on EULFS data.
$Y_{r,t}$	Gross domestic product	Gross domestic product (GDP) at current market prices per capita in region r at time t.	Authors' Calculation on Eurostat data and Regional Economy.
$E_{r,t}$	Tertiary education	Percentage of tertiary educated in the population in region r at time t.	Eurostat data.
$U_{r,t}$	Unemployment rate	Unemployment rate in region r at time t.	Eurostat data.
$P_{r,t}$	Population density	Total population over the area (km^2) in region r at time t.	Eurostat data.
$Soc_{r,t}$	Social benefits	Social benefits other than social transfer in kind per capita.	Authors' Calculation on Eurostat data
$Ch_{r,t}$	Children to adults ratio	Ratio of the total children aged 0 to 14 over the population aged 15 to 65	Eurostat data.

$C_{r,t}$	Total number of crimes	Total number of crimes, including robberies, homicides, burglaries and thefts in region r at time. t	Eurostat data.

Party

$Pro\ EU_p$	Pro EU political position	Measure party p pro EU political stance, computed as a difference between the shares of favorable and negative mentions in the political manifesto	Authors' calculation on MPD data
$Pro\ Nation._p$	Pro Nationalism political position	Measure party p pro Nationalism political stance, computed as a difference between the shares of favorable and negative mentions in the political manifesto	Authors' calculation on MPD data
$Total\ EU_p$	Salience EU issue	Measure of salience of EU topic for party p, computed as the sum of favorable and negative mentions in the political manifesto	Authors' calculation on MPD data
$Total\ Nation._p$	Salience Nationalism issue	Measure of salience of Nationalism topic for party p, computed as the sum of favorable and negative mentions in the political manifesto	Authors' calculation on MPD data

Table A6: Summary Statistics - Party closeness data

Party Closeness Data		Obs.	Mean	Std. Dev.	Min	Max
Individual Characteristics	Age	36155	51.593	18.082	18	90
	Female	36155	0.482	0.500	0	1
	Tertiary	36155	0.348	0.476	0	1
	Tertiary (father)	30501	0.212	0.409	0	1
	Preferences Pro EU	35749	0.014	0.031	-0.242	.080
	Salience EU	35749	0.033	0.019	0	0.262
	Preferences Pro National way	35749	0.012	0.020	-0.061	0.117
	Salience National Way	35749	0.015	.019	0	0.117
	Nationalism (PCA, std)	35749	0	1	-1.543	7.486
	Salience Nationalism (PCA)	35749	0.024	0.016	0	0.174
Regional Characteristics	GDP per capita	446	32.229	11.510	13.201	69.902
	Pop density	446	416.621	1025.116	3.3	7515.507
	Unemployment rate (%)	446	10.06	6.205	0	37
	Tertiary rate (%)	446	32.109	7.647	12.3	57.1
	Share of Mig (%)	446	13.28	8.837	1.7	50.7
	Share of Mig (HS) (%)	446	3.6	3.3	0.4	21.7
	Share of Mig (LS) (%)	446	9.7	6.0	1.3	34.7
	Share of Mig (not EU) (%)	446	7.2	5.2	0.6	32.4
	Share of Mig (not EU, HS) (%)	446	1.8	1.8	0	15.6
	Share of Mig (not EU, LS) (%)	446	5.4	3.7	0.4	23.2
	Share of Mig (EU) (%)	446	6.0	5.4	0	33.5
	Share of Mig (EU, HS) (%)	446	1.7	1.9	0	12.0
	Share of Mig (EU, LS) (%)	446	4.3	3.7	0	24.3

Note: authors' calculations on ESS data.

ESS voting

Table A7: Party closeness and party voted: differences (%)

Country	(1) Δ PC and PV	(2) DK Party close to	(3) DK Party voted	(4) Δ Tot. EU	(5) Δ Pro EU	(6) Δ Tot. Nation.	(7) Δ Pro Nation.
Austria	5.12	37.18	7.93	0.146	-0.054	0.096	0.063
Belgium	12.34	38.55	8.08	-0.044	-0.026	-0.035	0.015
Denmark	14.09	23.36	8.09	0.250	-0.486	0.495	0.397
Finland	11.69	31.93	11.43	-0.011	-0.103	0.032	0.032
France	30.00	39.84	16.47	-0.037	-0.026	-0.169	-0.193
Germany	15.27	32.71	11.32	0.071	-0.270	-0.001	-0.090
Ireland	28.32	55.43	6.25	0.625	-0.509	-0.582	-0.530
Portugal	9.16	22.36	16.90	0.062	-0.191	-0.023	-0.022
Spain	16.50	34.81	11.86	-0.259	-0.245	-0.016	-0.084
Sweden	14.06	24.62	6.32	-0.058	-0.257	-0.035	-0.024
Switzerland	25.14	15.94	24.88	0.382	-0.409	0.509	0.484
United Kingdom	12.30	33.53	11.52	-0.275	-0.098	0.021	-0.042
Whole Sample	16.76	34.01	12.33	0.103	-0.250	-0.007	-0.031

Note: authors' calculations on ESS data and Manifesto Project Database. Column (1) shows the share of individual who feels close to a party different from the party voted in the last national election. This share is computed only for the population who answered with a party name at both questions on party voted and party closeness. Column (2) shows the share of individual over the total population who answered "don't know" to the question related to the party close to. Column (3) shows the share of individual over the total population who answered "don't know" to the question related to the party voted in the last national election. From column (4) to (7) we compute the differences between the political platform of the party they feel close to and the party voted in the last national election. All the differences above are calculated among the population who feels close to a party different from the party voted in the last national election. The differences in the table are calculated as follow $\frac{(Party\ Close_j - Party\ Voted_j)}{sd(Party\ Voted_j)}$ for each topic j. All the differences in this tables are standardized by dividing the differences for the standard deviation of each topic among all the elected parties in the sample. Columns (4) and (6) show the std. differences on the total mentions of EU and National way of life topic respectively. Columns (5) and (7) show the std. differences on the favorable mentions of EU and Nationalism topic respectively.

Table A8: ESS data and actual voting (Δ)

Country	Year	(1) Top 5 parties Mean	(2) Top 5 parties SD	Country	Year	(1) Top 5 parties Mean	(2) Top 5 parties SD
Austria	2008	0.882	7.16	*Ireland*	2016	1.74	4.39
Austria	2013	0.831	5.16	*Portugal*	2009	0.972	7.13
Belgium	2010	1.61	2	*Portugal*	2011	1.52	8.18
Belgium	2014	1.36	1.61	*Portugal*	2015	-0.0255	9.7
Denmark	2007	0.17	2.96	*Spain*	2008	0.797	3.2
Denmark	2011	-.0823	1.78	*Spain*	2011	0.628	1.59
Finland	2007	0.16	3.09	*Sweden*	2010	0.428	3.39
Finland	2011	-0.304	3.44	*Sweden*	2014	0.004	3.87
Finland	2015	0.686	3.33	*Switzerland*	2007	0.0323	2.56
France	2007	0.381	4.71	*Switzerland*	2011	0.337	4.48
France	2012	1.45	3.39	*Switzerland*	2015	0.463	5.04
Germany	2009	0.876	7.88	*United Kingdom*	2010	-0.117	2.55
Germany	2013	1.42	4.22	*United Kingdom*	2015	0.106	3.35
Ireland	2011	0.7	7.78	**Whole Sample**	2007-2016	0.63	0.57

Note: authors' calculations on ESS, European Election Database (EED) and National Statistics. This table shows the mean and the standard deviation of the difference between voting shares computed with ESS and actual election results for the top 5 parties voted in each national election available in the sample.

3.9.3 Principal Component Analysis

We provide here below the results of the Principal Component Analysis on: (i)measures of favorable political position of the party, computed as the difference between positive and negative mentions, on the European Union and National Way of life; (ii)measures of saliency of European Union and National Way of life in the parties' political manifesto.

(i) Favorable political position

Table A9: PCA on favorable political position

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	1.51	1.03	0.758	0.758
Comp2	0.484	.	0.242	1.00

Table A10: PCA eigenvectors

Variable	Comp1	Comp2	Unexplained
Pro Nationalism	0.7071	0.7071	0
Pro EU	-0.7071	0.7071	0

(ii) Salience

Table A11: PCA on measures of salience

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	1.402	0.805	0.701	0.701
Comp2	0.597	.	0.298	1.00

Table A12: PCA eigenvectors

Variable	Comp1	Comp2	Unexplained
Total Nationalism	0.7071	0.7071	0
Total EU	0.7071	-0.7071	0

3.9.4 Parties' names and political preferences

Table A13: Partys' names and political preferences

Country	Party (abbrev)	Party (full name)	Nationalism	Salience
Austria	BZÖ	*Bündnis Zukunft Österreich (Alliance for the Future of Austria)*	0.681	0.023
	FPÖ	*Freiheitliche Partei Österreichs (Austrian Freedom Party)*	1.080	0.032
	TS	*Team Stronach für Österreich (Team Stronach for Austria)*	1.075	0.025
	NEOS	*Das Neue Österreich (The New Austria)*	-0.638	0.018
	LIF	*Liberales Forum (Liberal Forum)*	0.105	0.000
	GRÜNE	*Die Grünen (The Greens)*	-0.596	0.029
	SPÖ	*Sozialdemokratische Partei Österreichs (Austrian Social Democratic Party)*	-0.272	0.020
	KPÖ	*Kommunistische Partei Österreichs (Austrian Communist Party)*	0.819	0.033
	ÖVP	*Österreichische Volkspartei (Austrian People's Party)*	-0.258	0.029
	pirate	*Pirate Party*	0.200	0.006
Belgium	groen!	*Groen! (Green!)*	-0.215	0.008
	openVLD	*Open Vlaamse Liberalen en Demokraten (Open Flemish Liberals and Democrats)*	-0.294	0.011
	sp.a	*Socialistische Partij Anders (Socialist Party Different)*	-0.055	0.004
	ECOLO	*Écologistes Confédérés pour l'Organisation de Luttes Originales (Ecologists)*	-0.301	0.012
	CD&V	*Christen-Democratisch en Vlaams (Christian Democratic and Flemish)*	-0.565	0.016

	PSC	*Parti Social Chrétien (Christian Social Party)*	-0.376	0.012
	LDD	*Lijst Dedecker (List Dedecker)*	0.008	0.002
	N-VA	*Nieuw-Vlaamse Alliantie (New Flemish Alliance)*	-0.482	0.016
	VB	*Vlaams Belang (Flemish Interest)*	0.369	0.013
	PS	*Parti Socialiste (Francophone Socialist Party)*	-0.706	0.019
	MR	*Alliance: Mouvement Réformateur (Reform Movement)*	-0.418	0.015
Denmark	V	*Venstre (Liberals)*	-0.491	0.025
	SF	*Socialistisk Folkeparti (Socialist People's Party)*	0.398	0.013
	KrF	*Kristeligt Folkeparti (Christian People's Party)*	0.552	0.018
	SD	*Socialdemokratiet (Social Democratic Party)*	0.277	0.006
	KF	*Konservative Folkeparti (Conservative People's Party)*	-0.095	0.025
	EL	*Alliance: Enhedslisten - De Rød-Grønne (Red-Green Unity List)*	1.117	0.033
	DF	*Dansk Folkeparti (Danish People's Party)*	5.155	0.107
	NY	*Ny Alliance (New Alliance)*	0.105	0.021
	RV	*Det Radikale Venstre (Danish Social-Liberal Party)*	-1.028	0.024
Finland	SSDP	*Suomen Sosialidemokraattinen Puolue (Finnish Social Democrats)*	-0.043	0.006
	LKP	*Liberaalinen Kansanpuolue (Liberal People's Party)*	1.084	0.020
	VL	*Vihreä Liitto (Green Union)*	0.036	0.005
	PS	*Perussuomalaiset (True Finns)*	2.573	0.051

	VAS	*Vasemmistoliitto (Left Wing Alliance)*	-0.142	0.008
	SK	*Suomen Kansanpuolue (Finnish People's Party)*	-0.412	0.020
	pirate	*Pirate Party*	0.200	0.006
	RKP/SFP	*Ruotsalainen Kansanpuolue/Svenska Folkpartiet (Swedish People's Party)*	-0.348	0.009
	KD	*Suomen Kristillisdemokraatit (Christian Democrats in Finland)*	0.295	0.009
	KK	*Kansallinen Kokoomus (National Coalition)*	-0.273	0.024
France	NC	*Nouveau Centre (New Centre)*	-1.596	0.035
	Les Verts	*Les Verts, Confédération Écologiste - Parti Écologiste (The Greens)*	- 1.251	0.028
	FDG	*Alliance: Front de Gauche (Left Front)*	- 1.108	0.035
	UDF	*Union pour la Démocratie Française (Union for French Democracy)*	- 1.026	0.025
	PCF	*Parti Communiste Français (French Communist Party)*	- 0.047	0.020
	PRG	*Parti Radical de Gauche (Left Radical Party)*	- 1.520	0.039
	UMP	*Alliance: Union pour la Majorité Présidentielle (Union for the Presidential Majority)*	-0.355	0.054
	MoDem	*Mouvement Démocrate (Democratic Mouvement)*	- 0.620	0.047
	PS	*Parti Socialiste (Socialist Party)*	- 0.971	0.027
	FN	*Front National (National Front)*	3.533	0.073
	PR	*Parti Radical (Radical Party)*	- 0.580	0.066
Germany	LINKE	*Die Linke (The Left)*	-0.165	0.011
	SPD	*Sozialdemokratische Partei Deutschlands (Social Democratic Party of Germany)*	-0.366	0.021

	90/Greens	*Bündnis'90/Die Grünen (Alliance'90/Greens)*	-0.570	0.016
	CDU/CSU	*Alliance: Christlich-Demokratische Union/Christlich-Soziale Union (Christian Democratic Union/ Christian Social Union)*	0.297	0.033
	AfD	*Alternative für Deutschland (Alternative for Germany)*	4.181	0.090
	pirate	*Pirate Party*	0.200	0.006
	FDP	*Freie Demokratische Partei (Free Democratic Party)*	-0.593	0.025
Ireland	SoDeIR	*Social Democrats*	0.016	0.005
	irish	*Fine Gael (Familiy of the Irish)*	-0.198	0.020
	SF	*Sinn Féin (We Ourselves)*	1.375	0.032
	PBPA	*People Before Profit Alliance*	0.105	0.006
	ULA	*Alliance: United Left Alliance*	0.105	0.000
	Greens	*Green Party/Comhaontas Glas (Green Party)*	0.414	0.014
	AAA	*Anti-Austerity Alliance*	-0.219	0.015
	IndIR	*Independent Alliance*	1.446	0.041
	Labour	*Páirti Lucht Oibre (Labour Party)*	-0.128	0.013
	sd	*Fianna Fáil (Soldiers of Destiny)*	-0.045	0.019
Portugal	PCP	*Partido Comunista Português (Portuguese Communist Party)*	0.559	0.019
	CDS-PP	*Centro Democrático Social-Partido Popular (Social Democratic Center-Popular Party)*	0.192	0.020
	PS	*Partido Socialista (Socialist Party)*	-0.356	0.014
	PEV	*Partido Ecologista "Os Verdes" (Ecologist Party "The Greens")*	0.249	0.006

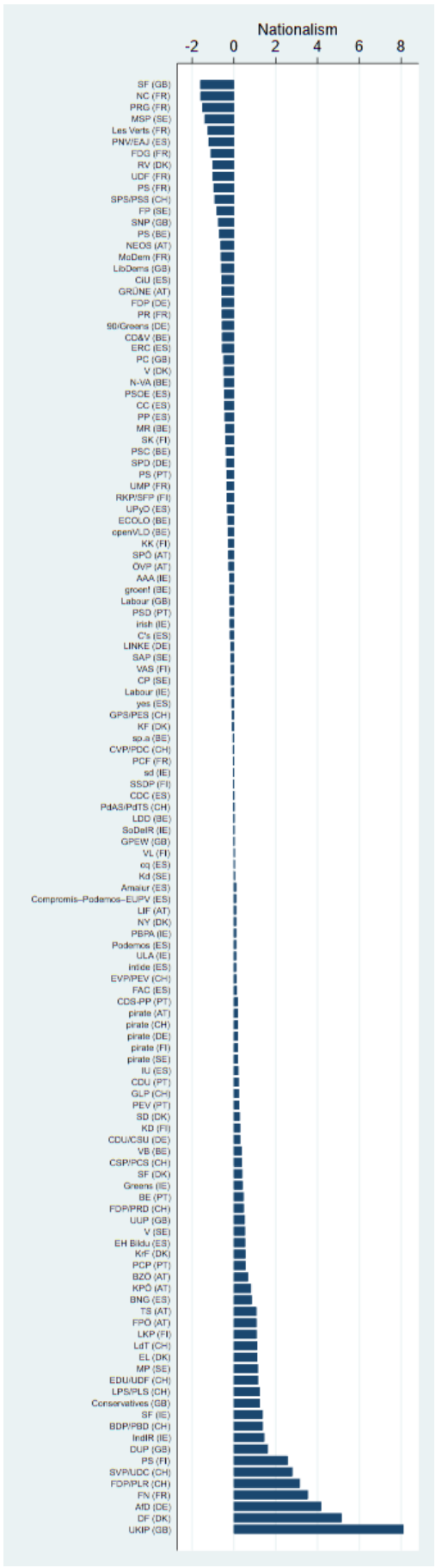
	PSD	*Partido Social Democrata (Social Democratic Party)*	-0.198	0.016
	CDU	*Coligação Democrática Unitária (Unified Democratic Coalition)*	0.226	0.017
	BE	*Bloco de Esquerda (Left Bloc)*	0.463	0.017
Spain	CC	*Coalición Canaria (Canarian Coalition)*	-0.468	0.013
	Amaiur	*Alliance: Amaiur (Amaiur)*	0.105	0.000
	CDC	*Convergència Democràtica de Catalunya (Democratic Convergence of Catalonia)*	-0.015	0.002
	EH Bildu	*Euskal Herria Bildu (Basque Country Unite)*	0.534	0.009
	intide	*En marea (In Tide)*	0.105	0.00
	C's	*Ciudadanos (Citizens)*	-0.191	0.013
	Podemos	*Unidos Podemos (United We Can)*	0.105	0.003
	BNG	*Bloque Nacionalista Galego (Galician Nationalist Bloc)*	0.849	0.020
	PNV/EAJ	*Partido Nacionalista Vasco/Euzko Alderdi Jeltzalea (Basque Nationalist Party)*	-1.215	0.032
	cq	*Alliance: Compromís-Q (Commitment-Q)*	0.039	0.001
	ERC	*Esquerra Republicana de Catalunya (Catalan Republican Left)*	-0.564	0.014
	CiU	*Alliance: Convergència i Unió (Convergence and Union)*	-0.598	0.015
	UPyD	*Unión, Progreso y Democracia (Union, Progress and Democracy)*	-0.323	0.009
	IU	*Izquierda Unida (United Left)*	0.213	0.010
	yes	*Alliance: Geroa Bai (Future Yes)*	-0.114	0.004
	PP	*Partido Popular (People's Party)*	-0.438	0.022

	PSOE	*Partido Socialista Obrero Español (Spanish Socialist Workers' Party)*	-0.477	0.017
	FAC	*Foro Asturias (Forum Asturias)*	0.141	0.011
Sweden	SAP	*Socialdemokratiska Arbetareparti (Social Democratic Labour Party)*	-0.156	0.016
	FP	*Folkpartiet Liberalerna (Liberal People's Party)*	-0.823	0.025
	MP	*Miljöpartiet de Gröna (Green Ecology Party)*	1.143	0.025
	Kd	*Kristdemokraterna (Christian Democrats)*	0.051	0.009
	pirate	*Pirate Party*	0.200	0.006
	CP	*Centerpartiet (Centre Party)*	-0.131	0.014
	V	*Vänsterpartiet (Left Party)*	0.527	0.014
	MSP	*Moderata Samlingspartiet (Moderate Coalition Party)*	-1.403	0.036
Switzerland	GLP	*Grünliberale Partei der Schweiz (Green Liberal Party)*	0.235	0.024
	FDP/PLR	*FDP.Die Liberalen/PLR.Les Libéraux-Radicaux (FDP.The Liberals)*	3.151	0.062
	GPS/PES	*Grüne Partei der Schweiz/Parti écologiste suisse (Green Party of Switzerland)*	-0.099	0.014
	SVP/UDC	*Schweizerische Volkspartei/Union démocratique du centre (Swiss People's Party)*	2.797	0.057
	SPS/PSS	*Sozialdemokratische Partei der Schweiz/Parti socialiste suisse (Social Democratic Party of Switzerland)*	-0.920	0.022

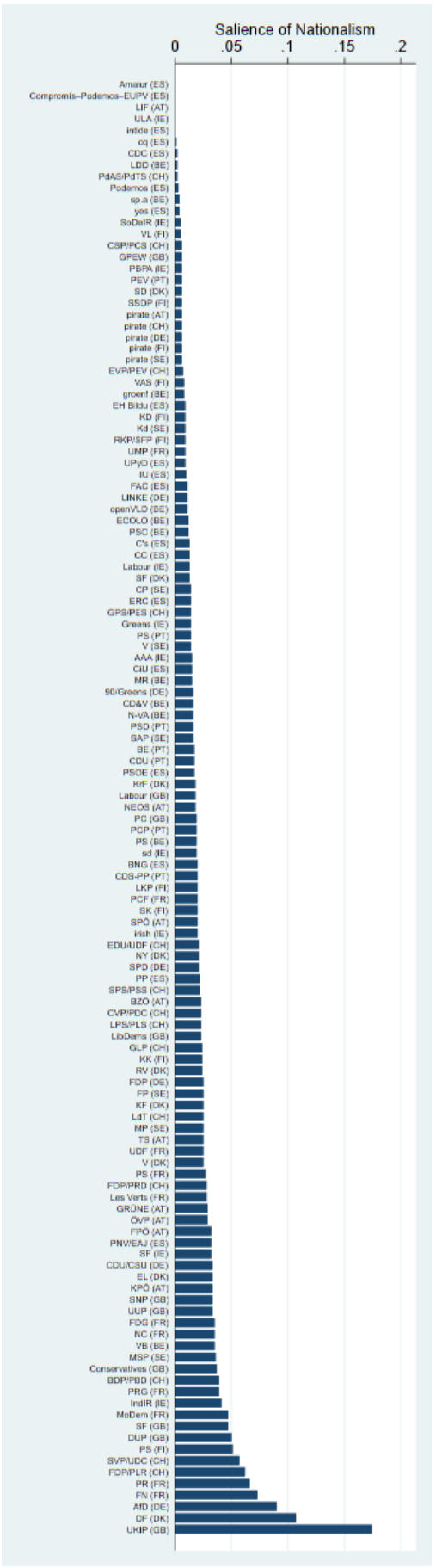
	EVP/PEV	*Evangelische Volkspartei der Schweiz/Parti Evangélique Suisse (Protestant People's Party of Switzerland)*	0.115	0.007
	BDP/PBD	*Bürgerlich-Demokratische Partei Schweiz/Parti Bourgeois (Conservative Democratic Party of Switzerland)*	1.377	0.039
	LPS/PLS	*Liberale Partei der Schweiz/Parti libéral suisse (Liberal Party of Switzerland)*	1.231	0.023
	CVP/PDC	*Parti démocrate-chrétien suisse (Christian Democratic People's Party of Switzerland)*	-0.053	0.023
	LdT	*Lega dei Ticinesi (Ticino League)*	1.116	0.025
	EDU/UDF	*Union Démocratique Fédérale (Federal Democratic Union)*	1.149	0.021
	CSP/PCS	*Christlich-soziale Partei/Parti Chrétien-Social (Christian Social Party)*	0.385	0.006
	FDP/PRD	*Parti radical-démocratique suisse (Radical Democratic Party)*	0.470	0.028
	PdAS/PdTS	*Partei der Arbeit der Schweiz/Parti suisse du travail (Swiss Labour Party)*	-0.006	0.002
	pirate	*Pirate Party*	0.200	0.006
United Kingdom	SNP	*Scottish National Party (Scottish National Party)*	-0.748	0.033
	UUP	*Ulster Unionist Party (Ulster Unionist Party)*	0.516	0.033
	LibDems	*Liberal Democrats (Liberal Democrats)*	-0.608	0.023
	PC	*Plaid Cymru (The Party of Wales)*	-0.495	0.019
	SF	*Sinn Féin (We Ourselves)*	-1.602	0.047
	Labour	*Labour Party (Labour Party)*	-0.204	0.018

Conservatives	*Conservative Party (Conservative Party)*	1.238	0.037
UKIP	*United Kingdom Independence Party (United Kingdom Independence Party)*	8.115	0.174
GPEW	*Green Party of England and Wales (Green Party of England and Wales)*	0.017	0.006
DUP	*Democratic Unionist Party (Democratic Unionist Party)*	1.620	0.050

Note: authors' calculations on ESS data and Manifesto Project Database. The table shows all the parties voted reported by the ESS and that won at least one seat in the parliament. The last two columns show the parties' political preferences measured with indicators of nationalism and Saliency.



(a) Nationalism



(b) Salience of Nationalism

Note: authors' calculations on Manifesto Project Database.

Figure A1: Nationalism, Salience and parties

3.9.5 Regional level Mechanisms

Table A14: Nationalism - Regional characteristics - Migrants Origin

	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	IV	IV	IV	IV
Time	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016
Regional Charact.	Social Transfer		Children		Crime	
Median	Below	Above	Below	Above	Below	Above
Not European Migrants						
Share HS	-0.51*	-0.04	0.05	-0.05	-0.20	-0.56**
	(0.27)	(0.36)	(0.29)	(0.95)	(0.17)	(0.25)
Share LS	0.19***	0.07	-0.02	-0.55	0.03	0.19***
	(0.07)	(0.11)	(0.03)	(0.93)	(0.05)	(0.07)
Observations	23633	24670	24871	23432	17438	30865
Regions	55	59	58	56	55	59
K-P rk Wald F-stat	10.87	0.80	0.49	0.23	3.43	10.10
Adj. R-Square	0.15	0.09	0.13	0.12	0.16	0.12
European Migrants						
Share HS	-0.82	0.18	0.46	-0.62	0.21***	0.51
	(1.19)	(0.12)	(0.46)	(1.37)	(0.08)	(5.56)
Share LS	0.93	-0.42**	-0.45	2.64	-0.24**	-2.14
	(2.08)	(0.17)	(0.44)	(15.55)	(0.09)	(12.68)
Observations	23633	24670	24871	23432	17438	30865
Regions	55	59	58	56	55	59
K-P rk Wald F-stat	0.11	5.09	0.27	0.01	4.12	0.01
Adj. R-Square	0.10	0.08	0.08	-0.01	0.16	-0.13
NUTS2 f.e.	✓	✓	✓	✓	✓	✓
Year f.e.	✓	✓	✓	✓	✓	✓
NUTS2 Controls	✓	✓	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓	✓	✓

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Standard errors are clustered at NUTS2 level. * p<0.1, ** p<0.05, *** p<0.01. As individual controls age, gender and education dummy are included. As NUTS2 controls we include log GDP per capita, unemployment rate and share of tertiary educated individuals. The dependent variable is our standardized measure of nationalism. Analysis by subsamples at regional level based on social trasfer per capita (col (1) and (2)), on the ratio $\frac{0-14}{15-65}$ old (col (3) and (4)) and on the total number of crime (col (5) and (6)).

3.9.6 Robustness Checks by regional economic conditions

Table A15: Nationalism - Subsample analysis based on unemployment rate quartiles

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	IV	IV	IV	IV	IV	IV	IV	IV
Unempl. Quartile	1^{st}	2^{nd}	3^{rd}	4^{th}	1^{st}	2^{nd}	3^{rd}	4^{th}
Time	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016
Share	0.00	0.09	-0.17***	-0.06				
	(0.04)	(0.09)	(0.02)	(0.05)				
Share HS					-0.11	-0.28***	-0.30***	-0.08
					(0.12)	(0.09)	(0.06)	(0.07)
Share LS					0.05	0.17**	-0.01	-0.03
					(0.10)	(0.07)	(0.05)	(0.05)
Observations	15741	10053	11792	10733	15741	10053	11792	10733
K-P rk Wald F-stat	12.61	37.53	135.25	1.18	2.52	9.84	8.72	0.82
Adj. R-Square	0.16	0.09	0.12	0.08	0.16	0.10	0.12	0.08
NUTS2 f.e.	✓	✓	✓	✓	✓	✓	✓	✓
Year f.e.	✓	✓	✓	✓	✓	✓	✓	✓
NUTS2 Controls	✓	✓	✓	✓	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Standard errors are clustered at NUTS2 level. * p<0.1, ** p<0.05, *** p<0.01. As individual controls age, gender and education dummy are included. As NUTS2 controls we include log GDP per capita, unemployment rate and share of tertiary educated individuals. The dependent variable is our standardized measure of nationalism. Quartiles are defined on the regional level of Unemployment rate in 2010.

Table A16: Nationalism - Subsample analysis based on Unemployment and Unemployment Variation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	IV	IV	IV	IV	IV	IV	IV	IV
Characteristics	Unemployment				Unemployment variation (2010-2007)			
Median	Below	Above	Below	Above	Below	Above	Below	Above
Time	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016
Share	0.05	-0.02			-0.10**	0.11**		
	(0.04)	(0.06)			(0.05)	(0.05)		
Share HS			-0.02	-0.19***			-0.24***	0.12
			(0.11)	(0.05)			(0.07)	(0.20)
Share LS			0.07*	0.05			0.01	0.11**
			(0.04)	(0.05)			(0.07)	(0.05)
Observations	27750	20569	27750	20569	26577	21742	26577	21742
K-P rk Wald F-stat	15.19	31.49	2.39	10.87	30.89	10.06	4.03	1.08
Adj. R-Square	0.16	0.10	0.16	0.10	0.16	0.09	0.16	0.09
NUTS2 f.e.	✓	✓	✓	✓	✓	✓	✓	✓
Year f.e.	✓	✓	✓	✓	✓	✓	✓	✓
NUTS2 Controls	✓	✓	✓	✓	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Standard errors are clustered at NUTS2 level. * p<0.1, ** p<0.05, *** p<0.01. As individual controls age, gender and education dummy are included. As NUTS2 controls we include log GDP per capita, unemployment rate and share of tertiary educated individuals. The dependent variable is our standardized measure of nationalism. Subsample analysis splitting regions above and below the country median level in 2010.

3.9.7 Vote for Party Families

Table A17: Voting Radical Right and Liberal parties - Voters Education

	(1)	(2)	(3)	(4)
	IV	IV	IV	IV
Time	2007-2016	2007-2016	2007-2016	2007-2016
Party Family	*Radical Right*		*Liberal*	
Natives Education	Not Tertiary	Tertiary	Not Tertiary	Tertiary
Share HS	-0.04*	-0.01	0.06*	0.01
	(0.02)	(0.01)	(0.04)	(0.04)
Share LS	0.02***	0.01**	0.00	0.01
	(0.01)	(0.00)	(0.01)	(0.01)
Observations	29575	16196	29575	16196
K-P rk Wald F-stat	4.70	10.58	4.70	10.58
Adj. R-Square	0.13	0.04	0.10	0.13
NUTS2 f.e.	✓	✓	✓	✓
Year f.e.	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓
NUTS2 Controls	✓	✓	✓	✓

Note: authors' calculations on ESS, EULFS, Manifesto Project Database, Chapell Hill Database and Eurostat data. Standard errors are clustered at NUTS2 level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$. As individual controls age, gender and education dummies are included. As NUTS2 controls we include the log GDP per capita, unemployment rate and share of tertiary educated individuals. The dependent variable is a dummy that take values of 1 if the individual voted for Radical Right parties in columns (1) and (2), while it takes value of 1 if the individual voted for Liberal parties in columns (3) and (4), following Chapell Hill codification. Data on Switzerland are not available.

3.9.8 Effects on Voting

Table A18: Voting - Voters political orientation

	(1)	(2)	(3)
	IV	IV	IV
Time	2007-2016	2007-2016	2007-2016
Subsample	*Political Orientation*		
	Left	Centre	Right
Share HS	-0.01	0.03*	-0.01
	(0.02)	(0.02)	(0.02)
Share LS	0.01	0.00	0.01
	(0.01)	(0.01)	(0.01)
Observations	16408	39339	17162
K-P rk Wald F-stat	21.26	20.46	14.55
Adj. R-Square	0.09	0.10	0.09
NUTS2 f.e.	✓	✓	✓
Year f.e.	✓	✓	✓
NUTS2 Controls	✓	✓	✓
Individual Controls	✓	✓	✓

Note: authors' calculations on ESS, EULFS and Eurostat data. Standard errors are clustered at NUTS2 level. * p<0.1, ** p<0.05, *** p<0.01. As individual controls age, gender and education dummies are included. As NUTS2 controls we include log GDP per capita, unemployment rate and share of tertiary educated individuals. The dependent variable is a dummy that take values of 1 if the individual voted during the last national elections. Analysis is run on subsamples based on the self-reported level of left-right political orientation of voters.

3.9.9 Nationalism and Migration attitudes

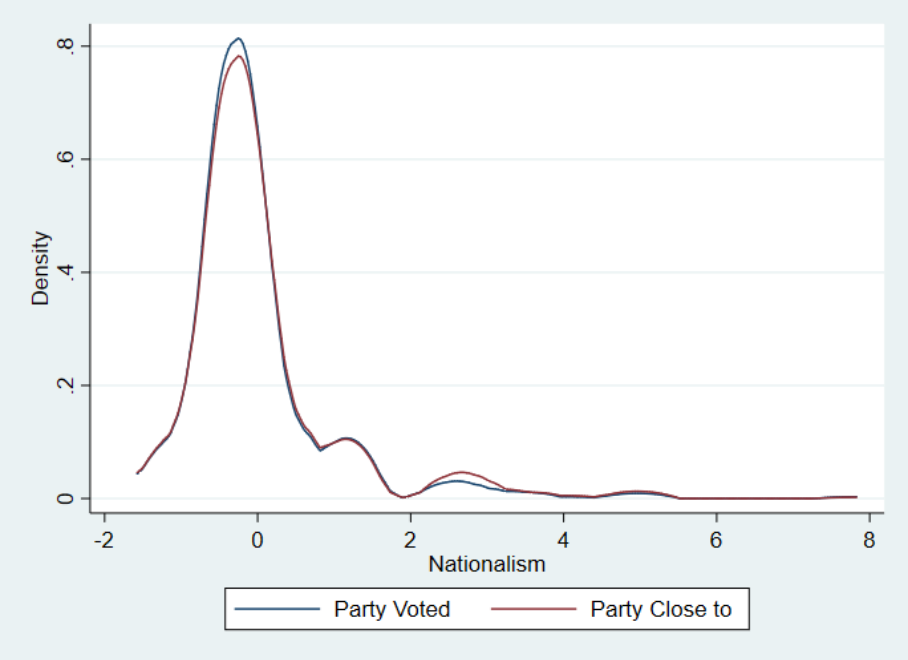
Table A19: Party Closeness - Nationalism and attitudes towards migration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	OLS	OLS	IV	OLS	OLS	IV	OLS	OLS	IV
Time	2010-2016	2010-2016	2010-2016	2010-2016	2010-2016	2010-2016	2010-2016	2010-2016	2010-2016
Share HS	-0.05**	-0.06**	-0.12**				-0.05**	-0.06**	-0.18
	(0.02)	(0.03)	(0.05)				(0.03)	(0.03)	(0.11)
Share LS				0.00	-0.00	-0.04	0.00	0.00	0.08
				(0.01)	(0.02)	(0.03)	(0.01)	(0.02)	(0.08)
Better Place to Live		-0.08***	-0.08***		-0.08***	-0.08***		-0.08***	-0.08***
		(0.01)	(0.01)		(0.01)	(0.01)		(0.01)	(0.01)
Enrich Culture		-0.14***	-0.14***		-0.14***	-0.14***		-0.14***	-0.14***
		(0.02)	(0.02)		(0.02)	(0.02)		(0.02)	(0.02)
Good Economy		-0.05***	-0.05***		-0.05***	-0.05***		-0.05***	-0.05***
		(0.01)	(0.01)		(0.01)	(0.01)		(0.01)	(0.01)
Observations	35121	35074	35074	35121	35074	35074	35121	35074	35074
K-P rk Wald F-stat			46.48			45.58			3.21
Adj. R-Square	0.14	0.18	0.18	0.14	0.18	0.18	0.14	0.18	0.18
NUTS2 f.e.	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year f.e.	✓	✓	✓	✓	✓	✓	✓	✓	✓
NUTS2 Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Standard errors are clustered at NUTS2 level. * p<0.1, ** p<0.05, *** p<0.01. As individual controls age, gender and education dummy are included. As NUTS2 controls we include log GDP per capita, unemployment rate and share of tertiary educated individuals. The dependent variable is our standardized measure of nationalism associated to the party which the respondent feel close to. As additional controls, we include standardized measures of respondent's attitudes towards migrants: whether migrants makes the country a better place to live (*Better Place to Live*), whether enrich country culture (*Enrich Culture*) and whether they have a positive effect on country economy (*Good Economy*).

3.9.10 Individual Panel Analysis

Figure A2: Nationalism Distribution



Note: authors' calculations on ESS data and Manifesto Project Database. The figure plot the distribution of the population in terms of nationalism.

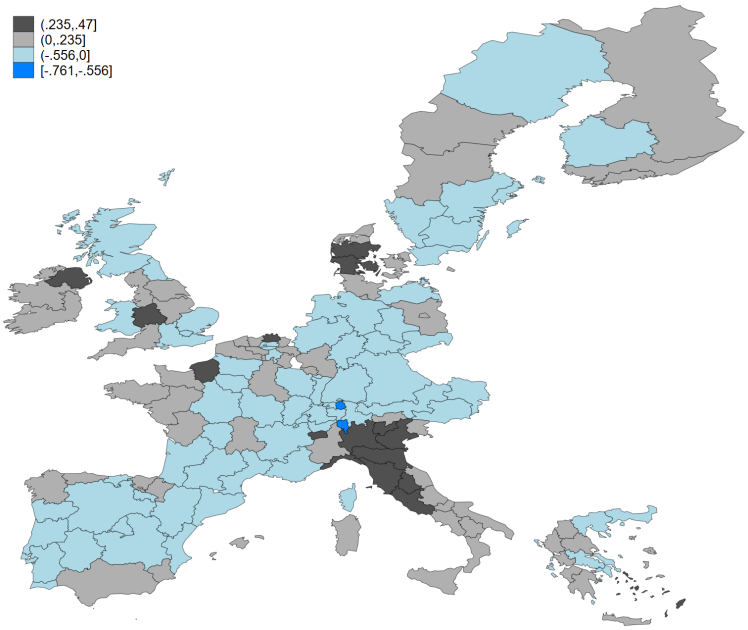
3.9.11 Simulations of different scenarios

Table A20: Simulations: exercises and robustness checks

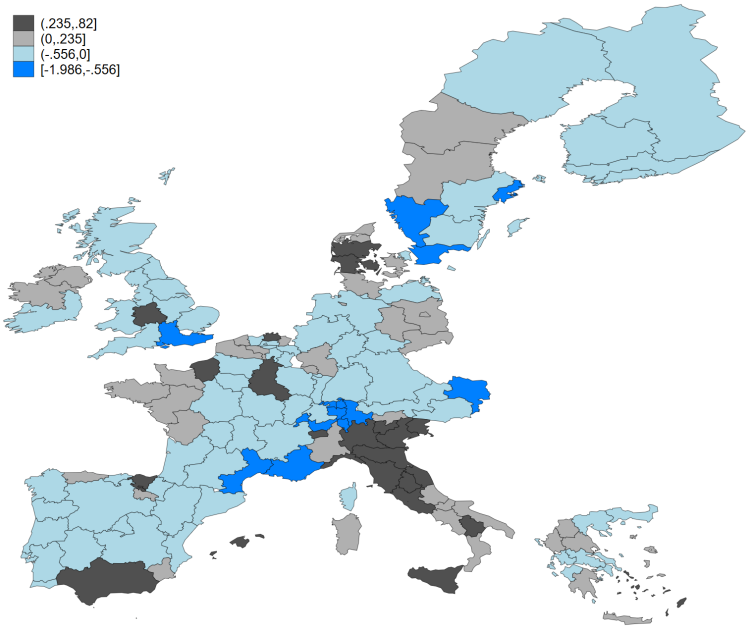
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Standard Sim.	Mig. Policies		Educ. Policies		Robustness Checks	
Country	$\widehat{Nation}_c$	*No Not EU LS*	*Balanced Not EU*	*Tertiary eff.*	*Max Tert. (EU)*	*Only LS Nat*	*Origin-specific*
Austria	-0.299	-0.364	-0.235	-0.36	-0.366	-0.241	-0.639
Belgium	0.042	-0.184	-0.161	-0.03	-0.094	0.06	-0.132
Denmark	0.171	0.035	-0.017	0.098	0.023	0.156	0.128
Finland	0.002	-0.082	-0.062	-0.08	-0.112	0.011	-0.09
France	-0.102	-0.171	-0.117	-0.169	-0.203	-0.075	-0.213
Germany	-0.051	-0.178	-0.166	-0.106	-0.162	-0.036	-0.127
Greece	-0.046	-0.029	-0.01	-0.104	-0.151	-0.037	-0.072
Ireland	0.085	-0.031	0.048	0.002	-0.041	0.076	-0.122
Italy	0.259	0.037	-0.001	0.224	0.059	0.246	0.313
Portugal	-0.116	-0.089	-0.021	-0.162	-0.202	-0.101	-0.254
Spain	-0.036	-0.029	-0.012	-0.105	-0.144	-0.027	-0.081
Sweden	-0.268	-0.535	-0.379	-0.347	-0.366	-0.181	-0.586
Switzerland	-0.508	-0.628	-0.489	-0.587	-0.581	-0.361	-0.986
United Kingdom	-0.1	-0.183	-0.069	-0.181	-0.223	-0.045	-0.466

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Column (1) shows country averages of the simulated standardized nationalism from the standard simulation using as regional weights the total population. Column (2) shows the result of the simulation when we remove the variation of low educated immigrants from not European countries. Column (3) shows the result of the simulation when we assume skill-balanced immigration from not European countries. Column (4) shows the result of the simulation when we include the direct effect of tertiary education. Column (5) shows the result of the simulation when each region has the same share of tertiary educated natives as the highest educated one in Europe (Great London). Column (6) shows the results of the simulation when we include in the equation only to low educated natives. Column (7) shows the results when we take into account country of origin specific effects.

Figure A3: Robustness Checks



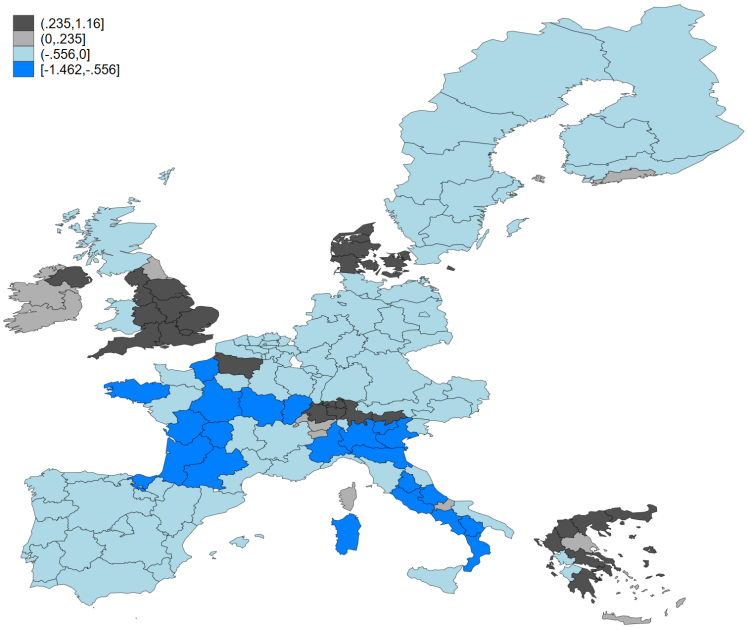
(a) Only LS Natives



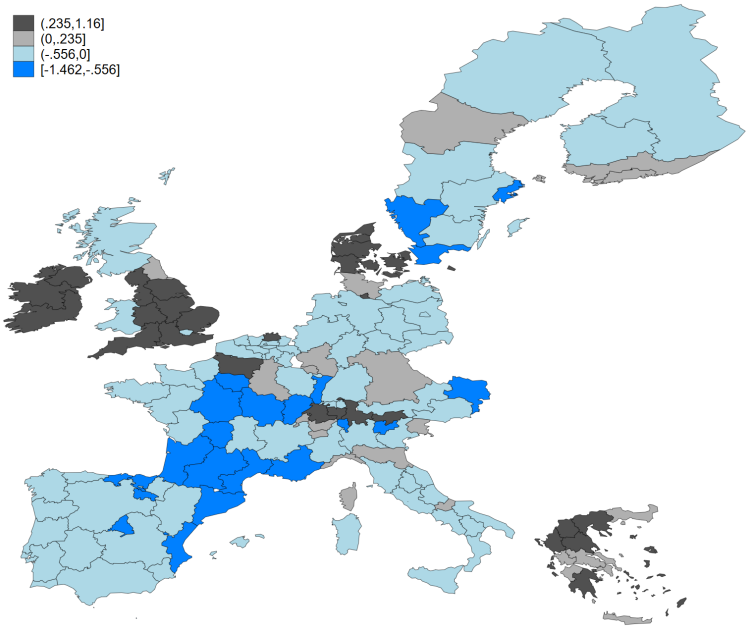
(b) Origin-skill specific analysis

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. The figure plots the simulated average level of nationalism after different scenarios, due to the variation of immigrants by education and level of education of natives over the 2007-2016. The figure plots the results of equation (3.10) at NUTS2 level when: only LS natives are included (panel (a)), skill and origin specific coefficients and migration variations are included (panel (b)).

Figure A4: Actual and predicted scenarios



(a) Results first election



(b) Results first election + sim.

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. The figure plots the average level of nationalism at NUTS2 level for the first election available in the sample (panel (a)) and how change the level of nationalism if we included the effect of immigration computed by equation (3.10) (panel (b)).

Table A21: Actual and predicted Nationalism

	(1)	(2)	(3)
Country	Year ($1^{st}elect$)	Nationalism ($1^{st}elect$)	Predicted Nationalism
Austria	2008	-0.116	-0.416
Belgium	2010	-0.396	-0.353
Denmark	2007	0.516	0.686
Finland	2007	-0.022	-0.019
France	2007	-0.486	-0.587
Germany	2009	-0.118	-0.170
Greece	2009	0.358	0.312
Ireland	2011	0.08	0.165
Italy	2013	-0.609	-0.351
Portugal	2009	-0.189	-0.306
Spain	2008	-0.435	-0.471
Sweden	2010	-0.469	-0.737
Switzerland	2007	0.588	0.08
United Kingdom	2010	0.327	0.228

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Column (1) shows the year of the first election available in our dataset. Column (2) shows the average level of Nationalism in the first election available, while column (3) shows the sum between column (2) and the variation on the level of nationalism from our standard simulation.

Chapter 4

Immigration and Voting for Redistribution: Evidence from European Elections

Abstract[1]

In this paper we document the impact of immigration on political support for welfare state expansion, using national election data of twelve European countries between 2007 and 2016. We match individual information on party voting with a classification of the political agenda of 126 parties during 28 elections. We first investigate the impact of local immigration on individual voting behavior, keeping the political platform of parties fixed. We then shift focus from voters to political parties, and investigate how immigration affects the political agenda of European parties. To attenuate omitted variable and selection bias concerns, we implement an instrumental variable approach that exploits cross-regional variation of immigrant settlements in 2005, along with the skill and nationality composition of recent immigrant flows. We find that larger inflows of *highly educated* immigrants are associated with European citizens shifting their votes toward parties that favor expansion of the welfare state. On the other hand, inflows of *less educated* immigrants induce European parties to endorse platforms less favourable to social welfare.

[1]This paper, coauthored with Giovanni Peri and Simone Moriconi, has been accepted by Labour Economics

4.1 Introduction

Economists have long argued that societies which are more ethnically and culturally diverse favour less redistribution and a lower supply of public goods (e.g. Alesina and Baqir and Easterly, 1999, Alesina and Gleaser and Sacerdote, 2001). This stems in part from an aversion to transferring income to groups ethnically or culturally different from one's own. More recently, some studies have found that immigrants may constitute a net fiscal cost for natives (see Orrenius, 2017), especially if their income is lower and if they choose their location based on the generosity of the local welfare state (the so-called welfare magnet theory). For both these reasons, the presence of a growing population of immigrants could make native citizens more inclined to downsize the welfare state and reduce fiscal redistribution and public spending.

Some evidence on the negative association between exposure to immigrants and support for welfare spending has been found in the sociological literature of the past decade (Eger, 2010, Senik *et al.*, 2009). Only recently, however, have some studies attempted to identify a causal impact of immigration on preferences for redistribution. Using census data to analyze the great migrations to the US in the early 20th century, Tabellini (2018) shows that increased immigration to US cities brought positive economic effects but also lower public support for the provision of public goods. Murard (2017) analyzes how the inflow of immigrants to European regions affected citizens' preferences for immigration and redistribution, using opinion surveys to measure the preference of citizens (rather than using their votes, as we do in this paper). It finds that an inflow of more-skilled immigrants produces stronger preferences for redistribution, while inflows of less-skilled immigrants weakens such preferences. Another recent paper by Alesina *et al.* (2018) analyzes cross-sectional variation, across countries and occupations, of both citizens' perceptions of immigrants and preferences for redistribution. The authors run an experiment in which people are randomly reminded of immigrants (or not) before answering questions about redistribution. They find that those individuals who are "primed" to think about immigrants are less likely to be in favor of redistribution. The paper also establishes that natives tend to overestimate the number of immigrants and their unemployment rates, suggesting that negative perceptions of immigrants contribute to weaker preferences for redistribution. Thus the paper not only directly addresses the connection between immigration and citizens' preferences for redistribution, but also suggests that perceptions of immigrants' characteristics contribute to the link. Neither Murard (2017) nor Alesina *et al.* (2018) analyze the party vote of individuals in response to immigration, which is the focus of our paper.

This paper tackles the following question: what is the impact of increased immigration on citizens' preferences for the welfare state, as revealed by their vote? The novelty of this paper is that we observe the party vote of citizens and that we systematically classify the political platforms of parties along the dimension of redistribution policies. Also new is the fact that we can decompose whether the change in policy preferences comes from a shift in vote shares across parties or from a change in parties' political platforms. An interesting and related question asks which features of immigrant populations generates stronger "aversion" to redistribution among citizens. Is the economic dimension important, such that less-educated and hence more "costly" immigrants, from a social welfare perspective, trigger a stronger aversion to the welfare state? (in line with the recent evidence from Murard (2017) and from Alesina *et al.* (2018b)). In our estimates, we focus on differentiating the impact of the presence of high- and less-skilled immigrants, given the fiscal impact of each type is quite different. We also check that these results are robust to considering only non-EU immigrants, whose inflow has grown particularly rapidly in the past decade.[2]

Finding an impact of immigrants on the vote of European citizens is interesting, as the result implies that immigration not only affects perceptions/preferences of citizens, but also that it is important (salient) enough to move people's vote. This is a novel contribution relative to existing papers, as it moves beyond the impact of immigration on individual opinions revealed by surveys (such as Hanson *et al.*, 2007 and Hangartner *et al.*, 2019). As political parties propose and support specific policies, an impact on voting behavior is certain to have implications on policies.

In a paper related to this one (Moriconi *et al.*, 2018), we use a similar method to analyze the impact of immigration on the vote of citizens for populist parties in European elections spanning 2007-2016. In that paper, the focus is on explaining whether the scale and the type of immigration experienced in the last decade can explain the electoral success of populist anti-immigration parties in many European countries during the recent electoral cycle (including France, Italy, Germany and the UK). We analyze how immigration flows are associated with votes for populist parties, and we find that only an inflow of less-skilled immigrants increases the propensity of natives to vote for populist parties, while an inflow of highly-skilled immigrants reduces that propensity. Recently, a large number of papers have analyzed the connection between immigration and vote shares for right wing (anti-immigration) parties. Most of that literature, however, differently from Moriconi *et al.* (2018), looks at party vote

[2]Chevalier *et al.* (2018) shows the importance of the votes of immigrants, themselves, in determining local redistribution, using the very special case of the repatriation of German citizens after World War II. In most cases, however, immigrants do not have voting rights; hence most of the literature focuses on their impact on electoral outcomes through their effect on the voting decisions (hereafter, the vote) of citizens.

(sometimes on a left-right scale) but not at specific implications for policies. Moreover, those studies use aggregate voting data, rather than individual voting. Notable contributions in this literature are Dustmann *et al.* (2018) who analyzed refugee resettlement and the vote in Denmark with a strong and credible identification strategy, Otto and Steinhardt (2014) who analyzed the effect of immigration on the vote for the German People's Union in Hamburg and Halla *et al.* (2017) who looked at votes to the Freedom Party of Austria.[3]

In this literature connecting immigration and citizens' voting decisions, the present paper is the first to focus on redistribution policies and to decompose the impact of immigration on the vote of individuals (hereafter, individual vote) and on party platforms relative to those policies. We use data from 12 countries in Europe and consider the period between 2007 and 2016, during which the immigrant share of the population rose significantly in all considered countries. We then analyze the individual vote as revealed by the data from the European Social Survey. The data include information on the party voted for, and we classify each party according to its stated position on net welfare transfers and on public education. Using text analysis, the Manifesto Project Database (MPD) produces for each party in Europe a measure of its support for generous welfare transfers. We standardize those indexes to have zero mean and standard deviation of one. Then we analyze the association between the regional inflow of immigrants and the change in the value of the index associated with an individual's vote. We separate immigrants between more- and less-skilled.[4] Connecting the individual party vote to the party position relative to redistribution policies is a novel feature of this paper, which allows us to translate a vote for a party into a vote for a specific policy, and to do this consistently across elections and countries.

Our analysis has two parts. First, we analyze whether the inflow of more- or less-educated immigrants is associated with a shift in individual vote towards or away from pro-redistribution parties (as revealed by their initial platform, measured by the MPD). In this analysis, we keep the platforms of parties fixed at their initial levels, and we measure changes in preferences for redistribution as changes in the individual votes received. While connected to the literature on the impact of immigration on the vote, the contribution of this part focuses on the implication for a specific policy. We find that while an increase in highly-skilled immigrants as a share of the population moves citizens to vote for pro-redistribution parties, an increase in

[3]Other papers have looked at several other countries. Barone *et al.* (2016) looked at immigration and vote to the Northern League coalition in Italy. Brunner and Kuhn (2018) analyzed the vote for the Swiss People's Party in Switzerland. Edo *et al.* (2019) looked at the French vote for the "Front National" in France. Mayda *et al.* (2018) showed that, in the US, high-skilled immigration has been strongly associated with a higher vote share for the Democratic Party, and less-skilled immigration with a higher vote share for the Republican Party.

[4]In a further specification, we focus only on immigrants from outside the EU.

less-skilled immigrants produces small and usually non-significant individual effects. Second, we analyze whether the same inflows of immigrants affected the policy platform of parties at the national level, considering a party-election year as one unit of observation. In this case, we analyze whether the position of a party, regarding redistribution, changes in response to immigration. This is a novel contribution of the paper, made possible by the time-varying classification of party preferences available in the MPD. We find that less-skilled immigration significantly shifts party platforms towards less redistribution while high-skilled immigration does not have a significant effect on a party's platform, although the point estimate of the impact is in favor of welfare state expansion.

Our results suggest that less-skilled immigration decreased native preferences for welfare state expansion, working through a change in party's platforms. Possibly as a response to the growing attention on immigration and its possible consequences, national parties adjusted their position by reducing their stated desires to expand the welfare state. On the other hand, high-skilled immigration moved individual native votes towards parties that are more in favor of welfare state expansions.

In order to reduce endogeneity and omitted variable concerns in our estimates, we instrument the inflow of immigrants by skill with the prediction obtained from their distribution across EU regions in 2005 and the aggregate inflow by nationality and skill level from 2005-2017. This is a version of the shift-share (enclave) instrument, which has been used in this literature for a long time (e.g. Altonji and Card, 1991, Card and Peri, 2016). In order to assuage worries arising from recent criticism of this instrument (e.g. Jaeger *et al.*, 2018) we test that this IV is not correlated with pre-2006 economic and demographic regional trends. The IV passes the validity test and the IV results generally confirm the OLS results in finding a pro-redistribution effect of high-skilled immigrants on natives' votes and an anti-redistribution effect of less- skilled immigration on party platforms.

Two additional findings emerge from our analysis. In analyzing the positions of parties, we also investigate their focus on public education. Spending in schooling may benefit immigrants and increase chances for economic and social integration. However, it can also be perceived as a transfer to immigrants, who on average have more children and who may require more resources (e.g. in language classes). Classifying parties based on their support for public education generates findings that are much weaker in response to immigration, and often not significant. Support for education may respond to a more complex set of issues, and more confounding factors can generate less-precise results. Finally, when performing a heterogeneity analysis, we find that the impact of high-skilled immigration in moving individual votes toward more pro-distribution parties is particularly strong for rural- and

less-educated voters.

The rest of the paper is organized as follows. Section 4.2 describes the data and the summary statistics. Section 4.3 presents the empirical specification and the identification strategy. Section 4.4 presents the main results on the citizens' vote and section 4.5 describes the results for parties' political platform. Section 4.6 concludes the paper.

4.2 Data and Summary Statistics

We combine data from multiple sources to create a dataset that includes the vote of a representative sample of citizens in each national elections, the party preferences for redistribution and public schooling (from their political manifestos), and the number of immigrants at the NUTS2 regional level for 12 EU-15 countries over the period 2007-2016.

We construct an indicator of the preference for redistribution of each political party in our sample based on its agenda (as described below), and link it to voters. Information on individual voters and their demographic characteristics are drawn from the *European Social Survey (ESS)*. This multi-country survey is administered once every two years.[5] On average, each wave contains around 1500 individuals for each country. This is meant to be a random sample of individuals that is representative of the population over 18 in each country. The data include detailed demographic information (i.e. age, gender, education), family status and characteristics (i.e. marital status, number of children in the family, place of birth), parental background (employment status, occupation when the respondent was 14 years old and country of birth for both parents) and labor market characteristics (e.g. employment status, occupation, type of contract).

ESS data also report the answer to a specific question on party voting behavior of individuals, namely: "*for which party did you vote in the last national election?*". Interviewees answer by providing the actual names of the parties voted in the country surveyed. We link the name of each party with the corresponding information on its political agenda, obtained from the *Manifesto Project Database (MPD)* (see Budge *et al.*, 2001; Klingemann et al., 2006). This database collects information on the policy positions of parties by using a content analysis of their electoral manifestos. The data cover over 1,000 parties, from 1945 until today, in over 50 countries on five continents. In this paper we focus on party preferences for redistribution. Based on MPD data, we measure preferences on this topic as the share of quasi-sentences related to two specific issues of redistribution over the total number of sentences in the party's

[5] As of 2019, there are 8 waves available. They cover 36 countries between 2002 and 2016.

political platform. The first policy we look at is each party's position on the expansion of the welfare state. In the definition of the welfare state, the MPD aggregates all mentions of health care, child care, care of the elderly, pensions and social housing.[6] It does not include public education (constituted by all types of education except for technical training), which is instead considered as a separate item and can therefore be analyzed on its own. We construct a measure of party preferences for *net welfare state expansion* as the difference in mentions between a welfare state expansion (variable per504 in MPD) and a welfare state limitation (per505). In a similar vein, we look at *net public education expansion* as the difference of mentions between a public education expansion (per506 in MPD) and a reduction (per507). These types of measures are typically used in the Political Science literature (see Budge *et al.*, 2001; Klingemann et al., 2006).[7]

In the first part of our analysis, we associate the value of the policy indicator for each party at the beginning of our period of analysis. In this case, variation in local preferences for redistribution derives only from the relative shift of voters from one party to another. In particular, we focus on within-region variation in voting behaviors between consecutive elections.[8] We aggregate individual data from the *European Labor Force Survey (EULFS)* to associate the stock of migrants in the region of residence in a given election year to each ESS respondent (and her vote). Besides demographic information, EULFS also reports the birthplace for each individual beginning in 2005, distinguishing fifteen different regions of origin.[9] We use all foreign-born individuals to compute a measure of immigrants, by skill-level, as a share of the total population in 2005 at the regional level, which is the key explanatory variable:

$$m^s_{r,t} = \frac{M^s_{r,t}}{Pop_{r,2005}}, \tag{4.1}$$

where $M^s_{r,t}$ is the total stock of migrants in region r born in a foreign country, with skills

[6]Because of this aggregation, it is not possible to analyze the position of parties on individual components of the welfare state such as health care or pensions.

[7]An alternative measure of salience of topics related to welfare and public education in a party agenda is the simple sum of positive and negative mentions. This indicator would capture any form of mention as increasing the salience of that theme for the party. Results based upon this indicator are available upon request.

[8]We use the intermediate geographic level, NUTS2, commonly indicated as "regions" in our analysis. ESS data report information on regional location of respondents within a country at the NUTS2 level for all EU countries with few exceptions (e.g. Austria, Germany, UK) where only NUTS1 location is consistently observed.

[9]These are the country-groups/regions of residence separately identified: EU15 country different from the country of residence, EU country that joined the EU in 2004, EU country that joined EU in 2007/2013, EFTA, Other Europan country, North Africa, Other Africa, Near and Middle East, East Asia, South and South East Asia, North America, Central America and Caribbean, South America and Australia and Oceania.

$s \in \{All, HS, LS\}$ in year t. Thus, $m^s_{r,t}$ is that group of immigrants as a share of the 2005 population.

The individual ESS data, which include the region of residence (NUTS2) and the party voted for, are linked to the EULFS data on immigrants in that region and to the MPD data on the political preferences (for redistribution) of the voted-for parties. We choose to focus only on core European countries, which are traditional net receivers of migrants. Thus we obtain our final sample, which includes information on 11 EU-15 countries plus Switzerland (these are Austria, Belgium, Denmark, Finland, France, Germany, Ireland, Portugal, Spain, Switzerland, Sweden and United Kingdom) for the 2007-2016 period. For each country, we build origin-skill-specific shares of migrants at the regional level.[10] The sample covers 114 regions, for which Eurostat and OECD data report regional variables which we use as controls in the analysis.[11]

Table 4.1: Elections and ESS rounds by country and year

	(1)	(2)	(3)	(4)
Country	# Elections	Election Years	# Survey Rounds	Survey Years
Austria	2	2008, 2013	3	2010, 2014, 2016
Belgium	2	2010, 2014	4	2010, 2012, 2014, 2016
Denmark	2	2007, 2011	3	2010, 2012, 2014
Finland	3	2007, 2011, 2015	4	2010, 2012, 2014, 2016
France	2	2007, 2012	4	2010, 2012, 2014, 2016
Germany	2	2009, 2013	4	2010, 2012, 2014, 2016
Ireland	2	2011, 2016	4	2010, 2012, 2014, 2016
Portugal	3	2009, 2011, 2015	4	2010, 2012, 2014, 2016
Spain	3	2008, 2011, 2016	4	2010, 2012, 2014, 2016
Sweden	2	2010, 2014	4	2010, 2012, 2014, 2016
Switzerland	3	2007, 2011, 2015	4	2010, 2012, 2014, 2016
United Kingdom	2	2010, 2015	4	2010, 2012, 2014, 2016

Note: Column (1) shows the number of elections available from ESS and column (2) the year of each elections. Column (3) shows the number of ESS rounds after 2010 by country and column (4) the year of each round. Source: ESS.

We only consider countries observed by the ESS for at least two national elections. We do this because our identifying variation comes from within-region changes in voting behavior

[10]Germany does not provide information on the birthplace of its foreign-born population. Accordingly, we impute the birthplace of the foreign-born population using information on the nationality of immigrants.

[11]These are GDP per capita, the unemployment rate, and the percentage of tertiary-educated individuals. See Table 4.2 for details.

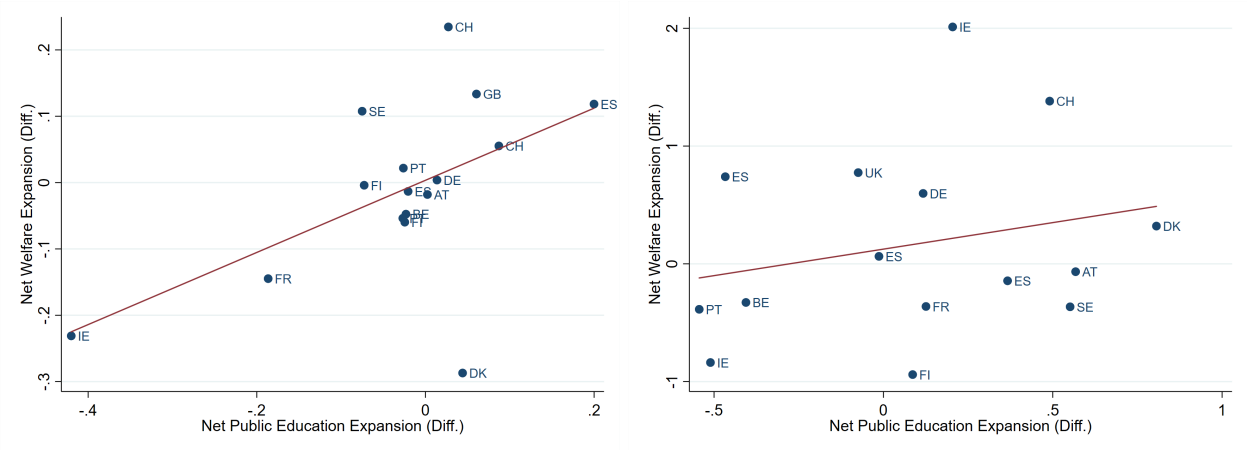
over different elections.[12] Table 4.1 summarizes the structure of our final dataset. It reports the number and years of elections covered by ESS questions for each country in columns (1) and (2), and the corresponding number and years of the surveys in columns (3) and (4). In the main part of the paper we look at individual votes in the latest elections, thus we exploit only changes in voting behaviors between elections. To make a practical example from Table 4.1, in the case of Sweden, we exploit variation in voting behaviors between election years 2010 and 2014 (see column (2)), even though the ESS survey was conducted in 2010, 2012, 2014 and 2016 (see column (4)). To do this, we compare the exact dates of the survey rounds and national elections. We are able to assign the vote of individuals surveyed in 2010 and 2012 to the 2010 elections because this is the most recent national election in which these respondents participated. Similarly, we assign responses to surveys in 2014 and 2016 to the 2014 Swedish national election if the date of the interview was after the 2014 election date, otherwise we assign the responses to the 2010 election. In this way, we map the variation between elections (column (2) in Table 4.1) into pseudo-panel variation of ESS data by survey years (column (4)).

In the baseline analysis, we explore the changes in individual votes for a party with a given political platform. We pool the individual data in repeated cross-sections and in this specification we keep the redistribution index of each party fixed at its initial level. We include region fixed effects in our analysis. We then perform an analysis exploiting changes in the political platform of parties, instead, looking at the national platform of each party and how its propensity for redistribution changed between elections in response to immigration in the country. The combination of individual-level and party-level analysis allows us a novel insight on how migration shaped preferences of a country's political system, through individual changes in party votes and through changes in party agendas.

Figure 4.1 shows a scatterplot of the change in preferences for the welfare state and support for public education expansion in EU countries between two consecutive elections in the period 2010-2016. In Panel (a), changes in political preferences at the country level are obtained by weighting political platforms of parties by the individual votes in the ESS dataset, keeping each party's platform fixed at the beginning of the period. In Panel (b), political-preference changes at the country level are derived from changes in the political platforms of parties and weighted using the average percentage of votes taken by each party over 2007-2016. Both panels show a positive correlation, indicating a relationship between the expansion of the welfare state and support for public education. In most countries–particularly Nordic and Continental European ones–political preferences shifted toward a more redistributive

[12]For this reason, we had to exclude Greece and Italy, for which ESS data only cover one election.

Figure 4.1: Political Preferences for Net Welfare and Public Education after Electoral Results



(a) Countries variation (individual voters)

(b) Countries variation (party platforms)

Note: authors' calculations on ESS and Manifesto Project Database. Figure (a) shows the average country level variation of political preferences for net welfare and public education expansion after each election event in our analysis where individual votes act as weights, keeping political platforms of parties fixed at the initial period. Figure (b) shows the country average variation of net welfare and public education expansion after each election event in our analysis using the average percentage of votes taken by each party over the 2007-2016 from MPD as weights and allowing parties political platform to change over time. The slope is statistically significant (1%) in Figure (a), while it is not statistically different from zero in Figure (b).

stance due to both individual vote shifting and to party platform changes (as both changes are positive in both panels).

The top panel of Table 4.2 presents descriptive statistics for our sample of individual voters. This includes over $50,000$ individuals between 18 and 90 years old who have the right to vote. The sample is balanced by sex. More than one third of these individuals have a tertiary education degree. On average, parties in our sample are highly favorable to redistribution policies: net mentions in the manifestos (i.e. the positive minus negative incidence of total mentions) are positive for expansions of both public education (5.4 on average) and public welfare (9.4 on average). The MPD indicator, featuring party-position on a hypothetical left-to-right political scale, shows that on average European voters tend to vote for parties with a relatively left-wing political manifesto.[13] This is consistent with the tendency to vote for parties that stress the role of the state as a provider of public services, a characteristic typical of left-wing parties. These indicators will be the main outcome variables of our empirical analysis. Given their different units, we standardize all of them so that a one-unit

[13]The position of each party on the left-right political scale is obtained by mapping policy preferences into a left vs. right score in each policy domain covered by MPD. Scores in all domains are then aggregated into a synthetic index that ranges between -100 (extreme left) to 100 (extreme right).

Table 4.2: Descriptive Statistics

Variable	(1) Obs	(2) Mean	(3) Std.Dev.	(4) Min	(5) Max
Individual Characteristics					
Age	50,862	52.501	16.972	18	90
Female	50,862	0.506	0.500	0	1
Tertiary	50,862	0.352	0.478	0	1
Net Welfare Exp.	50,304	9.373	6.471	-4.462	26.298
Net Pub. Edu. Exp.	50,304	5.441	3.244	0	16.609
Left-Right Index	50,304	-6.828	18.226	-48.000	72.500
Net Welfare Exp. (std)	50,304	0	1	-2.161	2.605
Net Pub. Edu. Exp. (std)	50,304	0	1	-1.680	3.443
Left-Right Index (std)	50,304	0	1	-2.259	4.352
Regional Characteristics					
GDP per capita	261	31.644	11.749	13.268	69.902
Pop. Density	261	420.592	1040.895	3.3	7393.4
Unemployment rate (%)	261	9.6326	5.492	0	30.1
Tertiary rate (%)	261	30.698	8.038	11.2	55.1
Share of Mig (%)	261	13.000	8.785	1.246	48.858
Share of Mig (HS) (%)	261	3.313	3.098	0	19.104
Share of Mig (LS) (%)	261	9.688	6.139	0.956	33.893
Share of Mig (not EU) (%)	261	7.115	5.020	0.197	30.568
Share of Mig (not EU, HS) (%)	261	1.690	1.626	0	13.356
Share of Mig (not EU, LS) (%)	261	5.4253	3.795	0	22.643
Share of Mig (EU) (%)	261	5.886	5.527	0	34.406
Share of Mig (EU, HS) (%)	261	1.623	1.869	0	11.730
Share of Mig (EU, LS) (%)	261	4.262	3.887	0	24.554
Parties Characteristics					
Received Votes (%)	233	10.952	11.010	0	45.24
Net Welfare Exp.	232	10.127	6.224	-4.93	28.169
Net Pub. Edu. Exp.	232	4.671	3.901	0	30
Net Welfare Exp. (std)	232	0	1	-1.197	6.493
Net Pub. Edu. Exp. (std)	232	0	1	-2.419	2.898

Note: authors' calculations on ESS, EULFS and MPD.

change corresponds to one standard deviation. This will allow meaningful comparisons of the estimated coefficients. The central panel of Table 4.2 shows the main summary statistics

of the main explanatory variable (i.e. migrants' population shares) and the regional controls (per capita GDP, population density, unemployment rate, and share of the population with tertiary-education). All of these variables vary at the regional NUTS2 level.

Immigrants accounted for 13% of the population of European regions during the 2007-2016 period. The majority came from outside the EU (7.1% of the population) and were less-skilled (9.7% of total population in the region). Finally, the bottom panel of Table 4.2 presents some descriptive statistics on the characteristics of the 233 parties in the sample. Each party received on average 11% of votes of European citizens. Averaging across parties we see that, on average, European parties are in favor of welfare expansion.

The EULFS (from which the immigration data are constructed) is representative of the European Population at the NUTS2 level, so our measure of immigrant share of the population should be unbiased. However, the ESS (hence the voting data) has not been explicitly designed to be representative at the NUTS2 level (though it is at the national level). As we use individual data from the ESS and we do not construct region-level statistics from it, this is less of an issue. Nevertheless, in Table 4.3 we compare the demographic composition of the sample of individuals in the ESS against the sample of the EULFS for the corresponding countries. We compare country- and NUTS2 level average shares of females, age groups, and shares of tertiary-educated individuals. When it comes to sex composition, the ESS and EULFS samples are not significantly different from each other (51% female, on average, both at the country- and the regional level). However, the ESS sample includes somewhat older individuals and a lower share of tertiary-educated, compared to EULFS. While statistically significant, such differences are not large, on average. Moreover, differences between ESS and EULFS averages are similar when we consider the country- and the NUTS2 level. As the two surveys are representative nationally, this shows that the sample differences do not arise due to regional variation. When interpreting the results, we keep in mind that older and less-educated individuals are slightly over-represented in our voting sample.

In Figure 4.2, for each country-election pair we plot vote shares calculated from the ESS data against actual vote shares available from the European Election Database (EED). This is a more direct test of how representative the ESS voting data are of total votes at the country level. In Panels (a) and (b) we focus on top-3 and top-5 parties in each country, respectively. Unsurprisingly, in countries characterized by proportional electoral systems (Belgium, Switzerland, and some Nordic countries), the total share of votes to top-3 parties tends to remain below 70% of total votes. More importantly, country-election vote shares are relatively close to the 45-degree line. This implies that the vote shares calculated using the ESS data and those calculated from the actual election data are quite similar. While the ESS

Table 4.3: Validation ESS data versus EULFS data

(1) Geo. Unit	(2) Variable	(3) Obs.	(4) $\overline{ESS}$	(5) $\overline{EULFS}$	(6)=(5)-(4) Diff.	(7) S.D.	(8) 95% Conf. Interval
Country	Female	72	0.509	0.511	0.002	0.034	[-0.007; 0.010]
Country	Age	72	47.64	46.72	-0.929***	2.600	[-1.541; -0.318]
Country	Tertiary Edu.	72	0.299	0.277	-0.022**	0.073	[-0.039; -0.004]
NUTS2	Female	648	0.506	0.511	0.005	0.125	[-0.004; 0.015]
NUTS2	Age	648	48.17	47.37	-0.809***	5.108	[-1.203; -0.415]
NUTS2	Tertiary Edu.	648	0.306	0.270	-0.036***	0.134	[-0.046; -0.026]

Note: this table shows the values of some socio-demographic characteristics at country and NUTS2 level using ESS and EULFS data. The respective means are presented in columns (4) and (5). The variables are the share of women in the population, the age and the share of tertiary educated individuals. Over the 2010-2016 period, the table shows the differences of demographic characteristics between EULFS and ESS data (column 6). Negative values imply higher values from ESS data compared to EULFS data. The stars represent the significance of the t-test for each difference being different from 0 (* p<0.1, ** p<0.05, *** p<0.01.). Column 7 show the standard deviation of the difference, while the lower and upper bound of the confidence interval are presented in column (8).

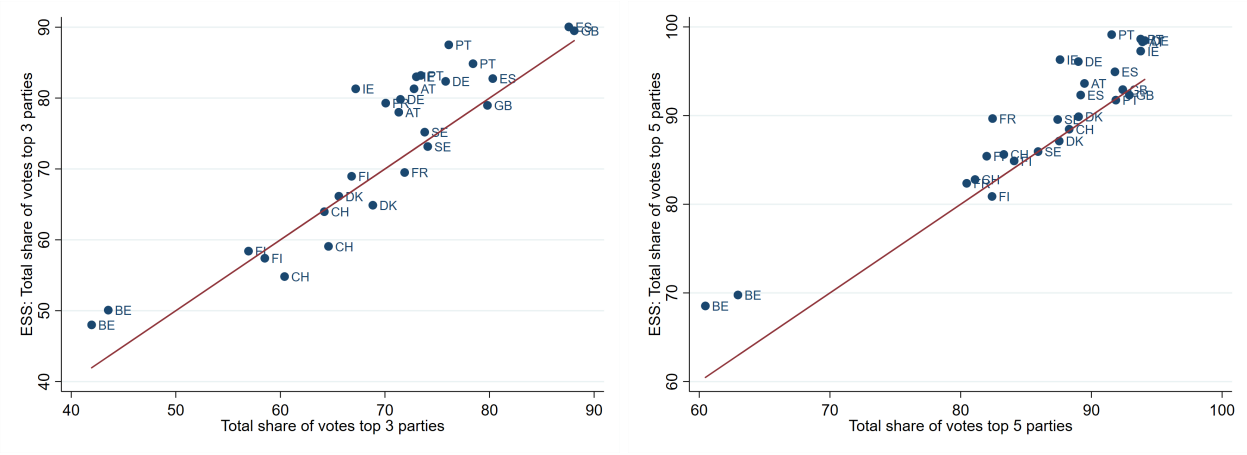
sample certainly has some measurement error, it does not seem to be large or systematically biased in one direction.[14]

Our empirical analysis looks at changes in voters preferences for state provision of welfare and education. We separate two mechanisms. First, we analyze the impact of immigration on the voting behavior of ESS voters, linked with parties' political agenda, keeping political agendas fixed (that is, we consider each political indicator for a party as fixed at the beginning of our sample). For parties that were born during our sample period, this would be their earliest available political platform. In the second part of the paper, we shift our focus to political parties and look at the impact of immigrants on changes in the political agendas of parties over time at the national level.

Figure 4.3 shows some features of these two complementary dimensions of the dataset. We aggregate parties' political platforms around party families, which represents some groups of parties that are ideologically similar, available from the *Chapel Hill dataset*. The Chapel Hill Expert Survey asked two hundred political experts to classify parties using the Hix and Lord (1997) classification. They also use information about the affiliation of national parties

[14]For a subset of NUTS2 regions and elections we can calculate the regional vote share of top-3 and top-5 parties from the ESS and from official election data. Figure A1, shown in the Appendix, confirms a strong proximity to the 45-degree line, with some measurement error.

Figure 4.2: Share of votes from ESS and Actual Share of Votes by Country Election in 2007-2016



(a) Top 3 parties

(b) Top 5 parties

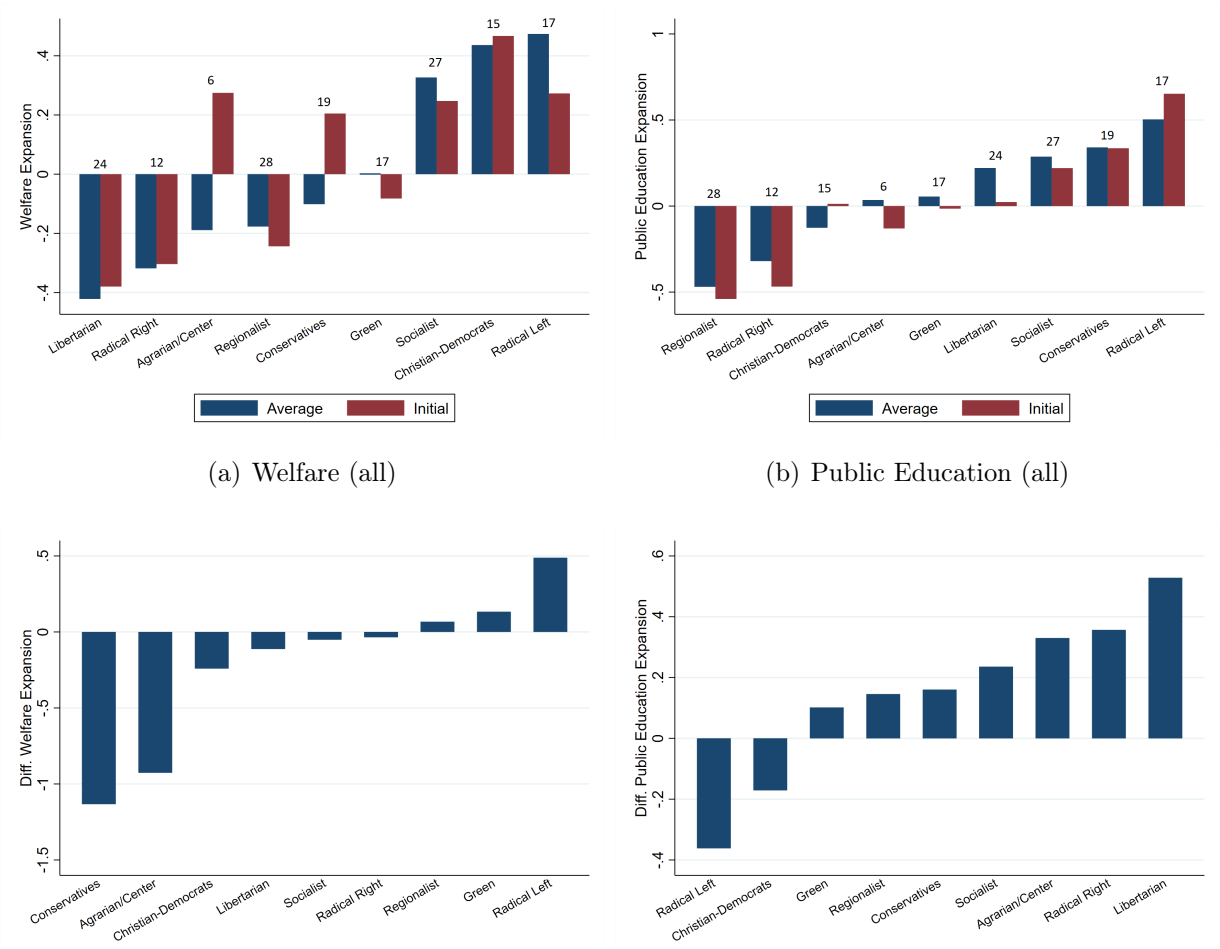
Note: authors' calculation on ESS and European Election Database (EED). The figure plots the share of votes of top 3 parties (panel (a)) and top 5 parties (panel (b)) in terms of votes in each election available computed with the EED (x-axis) and ESS (y-axis).

with international party associations to identify such political families. They identify eleven political families: Radical Right, Conservatives, Libertarian, Christian-Democrats, Socialist, Radical Left, Green, Regionalist, Confessional, Agrarian and a residual "No family" group. In the top panels, (a) and (b), we plot the average preferences in each family for welfare state- and public education expansions. The average of such preferences is standardized to zero and the standard deviation to one, and the graph shows, from left to right, families from the lowest to the highest preferences. We compute these preferences using both the average position of each party over the considered period (blue bar),[15] and the earliest average position of parties in the family (red bar), weighting each by the share of votes received by each party. On average, libertarian[16] and right-wing party families tend to have negative views regarding expansions of the welfare state and public education. Traditionally centre-left party families, instead, such as Socialists, Christian Democrats, and the Radical Left, have preferences for more redistribution. Generally speaking, average and initial platforms for a party family are strongly correlated. In many cases they do not differ too much. The variation in preferences over time for all party families is plotted in bottom panels, (c) and (d). Interestingly, only the Radical Left family has clearly shifted their political manifesto

[15]This is the average of each political indicator during the entire life span of the party itself.

[16]In the standard Chapel Hill classification this family is defined as "liberal", in the old-fashioned European sense of the word, which means believing in free markets. We redefine it as "Libertarian", which would be the modern American way of identifying this group.

Figure 4.3: Parties Families - Average position and Variation



(a) Welfare (all)

(b) Public Education (all)

(c) Welfare (Δ platform - balanced)

(d) Public Education(Δ platform - balanced)

Note: authors' calculations on Manifesto Project Database and Chapel Hill Database. Data on Switzerland are not available. Figures (a) and (b) plot the average position of political families, using the initial and the average political position of parties. Initial position is measured by the manifesto in the last available election before the start of the sample period, or the first year when the party is present in the political arena during the sample period (2007-2016). Average political position is measured as the average political manifesto of a party during the sample period 2007-2016. Number of parties in each family are reported over the bars. Figures (c) and (d) plot the variation of political families average position keeping a balanced sample of parties, i.e. including only parties that appear in multiple elections.

even further toward redistribution; on average, parties belonging to all other families have adopted platforms supporting a less-pervasive welfare state (panel (c)). At the same time, support for public education has increased for most party families, with the only exceptions being the Radical Left–whose support was and remains at the top of the spectrum–and Christian Democrat parties, whose change in this position was, however, quite small (panel (d)).

Overall, the panels of Figure 4.3 show that there are very important differences between parties in their positions towards public welfare and public schooling. Shifts of voters across parties can therefore change the political agenda of a country significantly in these areas. At the same time, some of the parties have changed their agenda on these issues. Understanding whether and how immigration has affected each of these changes is the goal of our empirical analysis.

4.3 Empirical Specification and Identification

In this section we present our empirical approach to investigating the impact of immigration on political preferences across EU regions. Subsection 4.3.1 focuses on the impact of immigration on the party preferences of EU voters by studying the change in preferences implied by the reallocation of voters across parties. Subsection 4.3.2, conversely, focuses on the impact of immigration on parties' political agendas, analyzing variation in the preferences of each party for public welfare and public schooling.

4.3.1 Impact on Voters' Preferences

Our benchmark empirical model is a simple panel regression that investigates the impact of immigration on voters' preferences for welfare and education spending, expressed through their vote. The dependent variable is the normalized indicator of welfare state expansion (or public education expansion) of the party voted for by citizen i in region r in election year e, ($Welf_{i,r,e}$ or $Edu_{i,r,e}$), as described in the previous section. In this section we assign to each voter the initial net political position of the voted-for party, defined by its manifesto as it was the first time the party participated in elections (within the 2006-2017 period). Hence the variation of preferences in region r over time that this measure captures is not driven by shifts in party agendas (which are kept fixed), but rather by a reallocation of voters across parties. The explanatory variable of interest is the adult population share of less- and more-educated immigrants, which proxies the exposure of citizens in a region to the presence of immigrants. Alternatively, to test whether the origin of immigrants matter , we also use the skill-specific population share of non-European adults as an explanatory variable. The sample of voters includes only natives, so we can interpret the coefficient as the impact of the local share of immigrants on the native vote. In our basic specification we include individual and regional characteristics as controls, with the goal of reducing unobserved variable bias

from multiple factors that may affect both immigrant presence and natives' votes. The basic specification we estimate is as follows:

$$Y_{i,r,e} = \alpha + \beta_H m^H_{r,e} + \beta_L m^L_{r,e} + \lambda \mathbf{X}_{i,r,e} + \gamma \mathbf{\Gamma}_{r,e} + \theta_e + \eta_r + \epsilon_{i,r,e}. \tag{4.2}$$

The dependent variables $Y = Welf, Edu$ are, respectively, the indicators of preferences for the welfare state and for public education of the party voted for by individual i. The two main explanatory variables of interest are $m^{H,o}_{r,e}$ and $m^{L,o}_{r,e}$. They represent migrants with tertiary education (H) or at most secondary education (L) as a share of the total 2005 population in region r and election year e. The vector $\mathbf{X}_{i,r,e}$ contains a set of voter characteristics including gender, age, and a dummy for tertiary education. The vector $\mathbf{\Gamma}_{r,e}$ includes a set of socio-economic variables at the regional (NUTS2) level, including the unemployment rate, GDP per capita (in logs) and the percentage of tertiary-educated individuals (in logs). Time invariant regional factors, like institutions, local culture, and geographic features, are captured by the regional fixed effects η_r. Common factors across elections are captured by election-year fixed effects θ_e. In an alternative specification, we replace our dependent variable with a synthetic indicator of the left-to-right position of the party voted by individual i in region r at election year e.[17]

Using least square methods to estimate equation (4.2) produces measures of partial correlation between the skill-origin-specific share of migrants and Europeans voting preferences, captured by the estimated coefficients $\hat{\beta}_H$ and $\hat{\beta}_L$. However, unobserved factors at the individual- and at the regional levels could remain in the error term and may be correlated with voters' preferences and immigrants' locations. In such cases, the estimated coefficients will be biased and will not reflect the causal effect of immigration on voting. We mitigate this potential bias by using an instrumental variable approach. Using a variation of the well-known shift-share methodology (see Altonji and Card, 1991, Card, 2001, and more recently Mayda *et al.*, 2018), we use past location of immigrants by country-of-origin as a predictor of subsequent inflows of immigrants. We consider previous immigrants' settlement, as measured by their shares in populations in 2005, to be uncorrelated (or at least not too correlated) with local factors that affected changes in voters' preferences in the 2007-2016 period. The changes in aggregate inflows of immigrants from different countries-of-origin–mainly driven by aggregate factors–are then allocated across EU regions following the early distribution of immigrants from the same origin-countries. Hence the instrument generates variation of regional migration over the period 2007-2016 by interacting aggregate national inflows

[17]Such an indicator, as described above, combines several aspects of the Manifesto in a synthetic measure of party location over the left-to-right political spectrum.

with previously established migration settlement patterns. Those components should be only weakly correlated with changes in political and economic characteristics at the regional level, reducing the severity of the omitted variable bias.

Since we want to capture the skill-origin specific effects of immigration, we construct two different instrumental variables, taking into account the aggregate variation of immigrant flows by skill and place-of-origin and applying it to the same initial distribution of immigrants by nationality across regions.[18] Namely, we define the initial presence of immigrants from origin country c in each NUTS2 region r as the share of their total population in EU12 in year 2005 as follows:

$$sh_{c,r,2005} = \frac{M_{c,r,2005}}{\sum_{r \in EU_{12}} M_{c,r,2005}}, \tag{4.3}$$

where $M_{c,r,2005}$ is the stock of immigrants born in place-of-origin c and residing in region r in year 2005, while the denominator is the total stock of immigrants born in place c and residing, as of 2005, in the aggregate of countries of our sample. The earliest year we can use to construct the regional distribution from the EULFS is 2005. The EULFS, in its harmonized codification scheme, grouped the origin countries into 15 groups that we use as "countries" of origin.[19] We then compute the total stocks of migrants resident in the EU12 countries, born in country-of-origin c in each year $t \in \{2007, 2008, ..., 2016\}$ separately for each of two skill levels s, equal to H (college educated) and L (at most a secondary education). That is, we compute:

$$TM^s_{c,t} = \sum_{r \in EU_{12}} M^s_{c,r,t} \text{ for } s = H, L. \tag{4.4}$$

We can then predict the total number (stock) of immigrants of skill s, from country-of-origin c in region r and year t as the interaction of those two values, as follows:

$$\widetilde{M}^s_{c,r,t} = sh_{c,r,2005} * TM^s_{c,t}. \tag{4.5}$$

We sum across all countries of origin and standardize for the 2005 total population in region r to have the predicted skilled and less-skilled immigrant shares of the population in region

[18] This is also done in Mayda *et al.* (2018).

[19] The fifteen birthplace regions recognized by the EULFS are the following: natives, EU15, New European member state from 2004, new European member state of 2007/2013, EFTA, Other Europe, North Africa, Other Africa, Near and Middle East, East Asia, South and South East Asia, North America, Central America and Caribbean, South America and Australia and Oceania.

r and year y as follows:

$$\widetilde{m^s}_{r,t} = \frac{\sum_c \widetilde{M}^s_{c,r,t}}{Pop_{r,2005}}. \tag{4.6}$$

We use these imputed shares as instrumental variables for the actual population shares of immigrants in each skill group H and L. Since we do not use any skill-specific pre-period distribution of migrants, the predicted power of the IVs will depend mainly on the tendency of individuals in each skill-origin group to locate themselves with previous immigrants from the same country-of-origin. The skill-specific instrument differs only due to the difference in skill-specific inflows from origin-countries.

Instrumental variables based on a shift-share methodology are widely used in the literature that looks at the causal relation between immigration and political outcomes (e.g. Otto and Steinhardt, 2014; Mendez and Cutillas, 2014; Barone *et al.*, 2016; Halla *et al.*, 2017; Brunner and Kuhn, 2018; Moriconi *et al.*, 2018). Nevertheless, recently, some articles have criticized these IV strategy, showing that existing and persistent correlation between economic/political conditions and the settlement of migrants may invalidate the IV (see Jaeger *et al.*, 2018 and Goldsmith-Pinkham *et al.*, 2018 for discussions). While we cannot test the exclusion restriction, we can check whether the IV is significantly correlated with pre-2007 trends in economic and demographic variables at the regional level which may affect voting behavior. The presence of such correlation would cast doubt on the validity of the instrument, as they could simply capture persistent regional trend that may affect votes.

In Table A1 we perform this validity check and show the correlation between the imputed growth of immigrants (total, high-, and less-skilled) in the 2007-2016 period and the pre-2007 trend of demographic and economic variables in NUTS2 regions. In particular, we regress the predicted growth of migrants (high- and less-skilled) as a share of the population on the growth of GDP per capita (top row), employment (second row), population (third row), and the share which are tertiary-educated (fourth row). Table A1 shows no significant correlation at the 1% level and only one coefficient out of 12 significant at the 10% level. Overall, the IVs for high- and less-skilled immigrants appear to be uncorrelated with pre-2007 economic and demographic trends.

Two further considerations suggest that the identifying variation driven by the 2007-2016 migration flows from important countries-of-origin was largely push-driven. First, the economic-financial crisis of 2007-2009 increased emigration, particularly of high-skilled individuals, from Southern and Eastern EU countries (see Anelli and Peri, 2017) and these people moved,

in large part, to Central and Northern European countries, generating flows driven in part by existing networks. Second, the accession to the European Union of 9 new member countries[20] between 2004 and 2006 produced large flows of migrants from those countries to Western Europe, driven by the freedom of movement granted to the new members (see Baas and Brücker, 2012). Both shocks pushed large numbers of immigrants in several North-Central and Western European countries, contributing to the variation in our IV and making this period one of particularly heavy migration into EU12 countries.

The estimation of equation (4.2) using the IV strategy described above provides coefficients that we cautiously interpret as the causal impact of immigrants on the average preference of a European voter. Different types of voters could react differently to the presence of immigrants. For example, Moriconi *et al.* (2018) shows that less-educated voters are particularly affected by inflows of immigrants, and Alesina *et al.* (2018b) shows that highly-educated Europeans are more open to redistribution when highly-educated immigrants join their countries. Studying the impact of immigration on Italian voters, Barone *et al.* (2016) shows that the significant effects are mainly driven by middle-sized municipalities. To explore the potential heterogeneity in the response across natives, we also estimate coefficients for sub-samples of voters. In particular, we split native population by education (tertiary and not-tertiary), age groups (18-37, 38-57, and 58+), location (rural areas, small towns, and cities and surroundings) and sex. The heterogeneity analysis provides a characterization of which demographic group responds to a greater extent to immigrants, and may suggest specific mechanisms at work in driving the vote of natives.

4.3.2 Impact on Parties' Platforms

Immigration may affect aggregate political preferences not only through voters' choice of party, but also by inducing parties to modify their political agendas. As the second innovative contribution of the paper, we investigate whether the inflow of immigrants drives parties to change their platforms on issues related to public education and redistribution. In this case, we define as dependent variables the net normalized preference for welfare expansion ($Welf_{p,d,e}$) and the net normalized preference for public education expansion ($Edu_{p,d,e}$) of party p in country d in election year e. The unit of observation is one party at the national level, and we exploit variation in each party's political manifesto between electoral events. The basic specification is the following:

[20]Signed on the 16^{th} of April 2003 by the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia and Slovenia, the Schengen Agreement became operative for these countries on December 2007.

$$Y_{p,d,e} = \alpha + \beta_H m^H_{d,e} + \beta_L m^L_{d,e} + \gamma \mathbf{\Omega}_{d,e} + \theta_e + \eta_d + \gamma_p + \epsilon_{p,d,e} \tag{4.7}$$

where $Y = Welf, Edu$. Our main variables of interest are $m^H_{d,e}$ and $m^L_{d,e}$, which represent the share of migrants with high- (tertiary) and less- (secondary and less) education over the total 2005 population in country-of-destination d in election year e. The vector $\mathbf{\Omega}_{d,e}$ contains a set of country-level demographic and economic characteristics, reflecting the controls used at the regional level. They include the (log) share of tertiary-educated individuals, (log) GDP per capita, and the unemployment rate. The presence of party fixed effects (γ_p) and country fixed effects (η_d) implies that we identify the effect of immigration on political agendas only through changes in each variable over time. Any time-invariant feature of countries and parties will not affect identification, as it will be absorbed by these fixed effects. Finally, election-year fixed effects (θ_e) capture common trends over time.

We adopt the same identification strategy discussed above by constructing the imputed level of high- and less-skilled immigrants, this time at the country level. In particular, we allocate immigrants during the sample period based on their distribution across European countries in 2005 by country-of-origin. Even in this case, the goal of using this shift-share approach is to reduce the correlation of immigration flows with unobservable contemporary political and economic changes which may affect the vote. We construct the indicators (4.3), (4.4) and (4.5), and we modify equation (4.6) to match the national unit of analysis as follows:

$$\widetilde{m}^s_{d,t} = \frac{\sum_{r \in d} \sum_c \widetilde{M}^s_{c,r,t}}{Pop_{d,2005}}, \tag{4.8}$$

Expression (4.8) adds the predicted regional stocks of immigrants across all regions in a country, and predicts the skill-specific shares of immigrants in the 2005 population in destination country d and year t. Using the predicted immigrants shares at country level as IVs, we estimate our equations both with Least Squares and 2SLS.

In the baseline results, we give each party the same weight so as to capture the average effect of immigration on the party system's political preferences. However, small parties do not have the same influence on a country's political system as large parties, and they may change their positions more easily. Giving equal weighting to small and large parties may overestimate the changes in platforms. For this reason, we perform some robustness exercises. First, we focus on the sample of parties with at least 10% of the votes in the elections. Exclusion of small parties considerably reduced sample size (from 74 to 43 parties). Second, we weight each party by the percentage of votes gained in the last elections. Taken together, these

results provide a more complete understanding of the effect of immigration on the average agenda of political parties in a country.

Note that, because of the inclusion of party- and country fixed effects, the impact of immigration on parties' political preferences estimated from equation (4.7) are only identified by changes in agendas for parties which were present in at least 2 elections. This captures genuine changes in the agendas of *existing* parties in response to changes in immigrant shares, rather than the entry or exit of *new* parties.

4.4 Results on Voters' Preferences

Our first set of results is relative to the impact of immigration on the intensity of political preferences for welfare state- and public education expansion. In Panel A of Table 4.4, we report the estimates of interest from specification (4.2), where we use party preference for the welfare state as the dependent variable. Panel B shows the estimates when the dependent variable is a party's net preference for public education expansion. We run different specifications using OLS (odd number columns) or IV (even number columns) for estimation. In different rows, we either include total immigrants or high- and less-skilled immigrants, separately, as explanatory variables, while each panel includes all immigrants, in the upper part, or only non-EU immigrants, in the lower part, to calculate the explanatory variable. In columns (1)-(2) we use the share of total immigrants relative to the 2005 population as the main explanatory variable. In columns (3)-(4) and (5)-(6) we focus on the shares of highly-educated and less-educated migrants, respectively. Finally, in columns (7)-(8) we use both high- and less-skilled migrant shares, together, as explanatory variables.

All regressions include regional and individual controls plus election year- and region (NUTS2) fixed-effects. Standard errors are also clustered at the regional level, allowing correlation over time in the unobserved component of regional votes. The power of our instrumental variables is reasonably strong. When included one at a time, the predicted share of high- and less-skilled immigrants has an F-stat between 37 and 56, which is quite large. When included together, the two IVs still have an F-statistic of 15. Concerns of weak instrument bias should be reasonably minor. OLS and IV produce qualitatively similar results, with slightly larger coefficient estimates using the IV method.

Baseline estimates in Panel A of Table 4.4 suggest that an increase in the average share of immigrants pushes voters towards parties with higher propensities for welfare state expansion. However, when we distinguish immigrants by skill-group, we find that the effect is

Table 4.4: Political Preferences over 2007-2016 - Individual Voters

	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
Panel A - Net Welfare State								
Share	0.007	0.014*						
	(0.006)	(0.008)						
Share HS			0.030**	0.049**			0.033**	0.052**
			(0.013)	(0.022)			(0.015)	(0.022)
Share LS					0.001	0.009	-0.007	-0.007
					(0.007)	(0.012)	(0.007)	(0.011)
Observations	50304	50304	50304	50304	50304	50304	50304	50304
K-P rk Wald F-stat		48.36		32.87		32.58		15.12
Adj. R-Square	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70
Share No EU	0.002	0.004						
	(0.007)	(0.010)						
Share No EU HS			0.020	0.077*			0.023	0.105*
			(0.025)	(0.045)			(0.026)	(0.055)
Share No EU LS					-0.001	-0.002	-0.005	-0.028
					(0.008)	(0.013)	(0.008)	(0.018)
Observations	50304	50304	50304	50304	50304	50304	50304	50304
K-P rk Wald F-stat		38.63		55.30		37.03		16.45
Adj. R-Square	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70
Panel B - Net Public Education								
Share	-0.004	-0.007						
	(0.006)	(0.016)						
Share HS			0.007	0.058***			0.014	0.080***
			(0.011)	(0.020)			(0.013)	(0.026)
Share LS					-0.010	-0.028	-0.014*	-0.053**
					(0.007)	(0.022)	(0.008)	(0.021)
Observations	50304	50304	50304	50304	50304	50304	50304	50304
K-P rk Wald F-stat		48.36		32.87		32.58		15.12
Adj. R-Square	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56
Share No EU	-0.017**	-0.010						
	(0.008)	(0.018)						
Share No EU HS			-0.020	0.133***			-0.006	0.214***
			(0.026)	(0.050)			(0.027)	(0.071)
Share No EU LS					-0.021**	-0.028	-0.020**	-0.080***
					(0.009)	(0.020)	(0.009)	(0.026)
Observations	50304	50304	50304	50304	50304	50304	50304	50304
K-P rk Wald F-stat		38.63		55.30		37.03		16.45
Adj. R-Square	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56
NUTS2 f.e.	✓	✓	✓	✓	✓	✓	✓	✓
Election f.e.	✓	✓	✓	✓	✓	✓	✓	✓
NUTS2 Controls	✓	✓	✓	✓	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Standard errors are clustered at NUTS2 level. * p<0.1, ** p<0.05, *** p<0.01. As individual controls age, gender and education dummies are included. As NUTS2 controls we include the logarithm of GDP per capita, of the share of tertiary educated individuals and the unemployment rate. The dependent variables are the standardized net favorable position towards welfare expansion (Panel A) and net favorable position towards public education expansion (Panel B) using the initial position of parties over our time span.

driven by the share of high-skilled migrants. The coefficient on highly-educated immigrants as a share of the population estimated using IV (column 4) is positive and significant, while the share of less-educated immigrants has a non-significant coefficient (column 6). These results are confirmed when we include both high and less-skilled immigrant shares at the same time (column 8). The difference in the estimates between OLS and IV is small, with IV estimates larger than OLS. This suggests a negative correlation between the error and the explanatory variable. This may happen, for instance if regions where natives have moved towards pro-redistribution political platforms (e.g. in response to bad economic outcomes) are also less likely to attract high-skilled immigrants. We evaluate the magnitudes of the effects using the estimated coefficients in column (8). An increase of one standard deviation in the share of highly educated immigrants (equal to 3.09 percent of the population) will shift voters toward a more redistributive agenda by 0.16(=3.09 * 0.052) standard deviations of the net welfare expansion indicator. The results associated with non-European migrants, shown in the lower part of Panel A, are in line with (and have a relatively larger size than) our baseline results.[21] The pro-redistribution effects of highly educated immigrants are somewhat stronger in the presence of an increase of non-Europeans immigrants. The coefficients, however, are less precisely estimated.

Panel B of Table 4.4 shows the results using the net party preference for public education expansion as a dependent variable. The average effect of the share of immigrants on voting behavior towards public education expansion is not statistically significant. Even in this case, however, when we separate immigrant shares by education levels, we find results similar to those on voting preferences towards the welfare state. An increase of the share of highly-educated immigrants is associated with voters shifting towards parties that want to expand public education. A higher share of less-educated immigrants instead pushes voters toward parties that are less-interested in investing in public education. The results for non-European immigrants are even stronger with a larger difference in the positive effect of non-European high-skilled and the negative effect of non-European less-skilled immigrants. Education is often regarded as a potential channel of inclusion and assimilation of newly arrived migrants, in particular if they are poorly educated. However, natives may not be willing to invest in local public goods and services, such as public schools, fearing that these will benefit primarily less-skilled (and low-income) immigrants (Card, 2012). Using the value of the coefficients in column (8), an increase of one standard deviation of highly educated immigrants is associated with a shift of voters toward parties with an agenda 0.25(=3.09*0.080) standard deviations

[21] We do not report the estimates for European immigrants only as the instrument is quite weak in that case. They are available upon request. As the IV variation is mainly driven by non-EU migrants, we can say that the overall LATE estimated by the 2SLS is potentially driven by non-EU-immigrant variation.

more favorable to net expansion of public education. Instead, the effect of an increase in the share of less-educated immigrants, obtained multiplying one standard deviation of that variable (6.14) times the coefficient of column (8), -0.053, implies a change by 0.32 standard deviations towards parties less-favourable to public schools. The negative effect is even stronger when it is associated with non-European, less-educated immigrants.

We perform two checks of these results. First, in the specifications in Table 4.4, we keep the position of a party, relative to the two policies considered, as it was when the party first entered the political scene (within our period of interest). If we use the average position over the 2006-2017 period, the results are very similar, as the results of Table A2 in the Appendix show. Second, in Table A3 of the Appendix we replicate the analysis using as dependent variable a normalized indicator on the left-to-right position of voters. In this way we test the effect of immigration on a broader indicator of a party's position on the political spectrum. Higher values of the indicator are associated with a party being more right-wing. In-line with previous results, an increase in highly-educated immigration is associated with a shift of voters toward more left-wing parties (negative coefficient). Less-skilled immigrants have negative to zero effects. Consistent with the above findings, left-leaning parties are, on average, more likely to support an inclusive and larger welfare system.

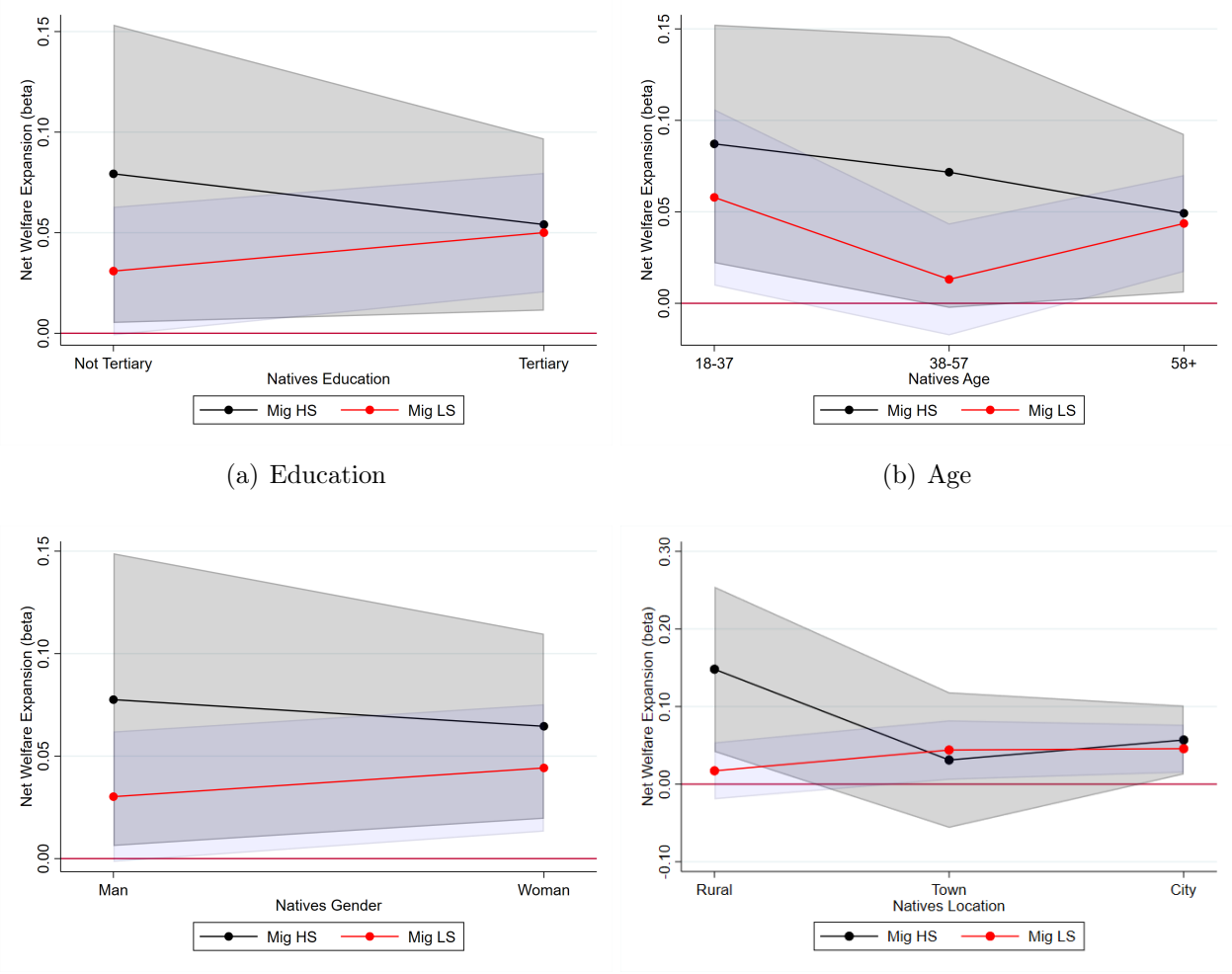
4.4.1 Heterogeneity in Individuals' Voting Response

The results presented in the previous section show the average effect of high- and less-skilled immigrants on citizens' political preferences. However, these effects can be heterogeneous across individuals. To explore this issue, we perform an analysis of different sub-samples of voters based on individual characteristics. We split the native population by education level between those with tertiary- and those without tertiary education. We split these again into three age groups: 18-37, 38-57, and 58+. We split by location in rural areas, small towns, and urban areas, and by sex. We estimate the effects on voting by including the skill-specific immigration shares separately, and by instrumenting them with the shift-share instruments. The point estimates and confidence intervals of the effects on the net welfare state expansion are plotted in Figure 4.4, while the effects on net public education expansion are plotted in Figure 4.5.[22] Shaded areas denote the 95% confidence intervals around the estimated value. The black line connects the coefficient values in response to high-skilled immigration (Mig HS). The red line connects the estimates obtained in response to less-skilled migration (Mig

[22]The Tables with the corresponding estimated values are included in the Appendix. Table A4 shows the subsample analysis related to net welfare state expansion, while Table A5 shows the subsample analysis for net public education expansion.

LS). The average F-statistic associated with our instruments is above 10, showing that our IVs have enough explanatory power.

Figure 4.4: Net Welfare State Expansion and Natives' Individual Characteristics



(a) Education

(b) Age

(c) Gender

(d) Location

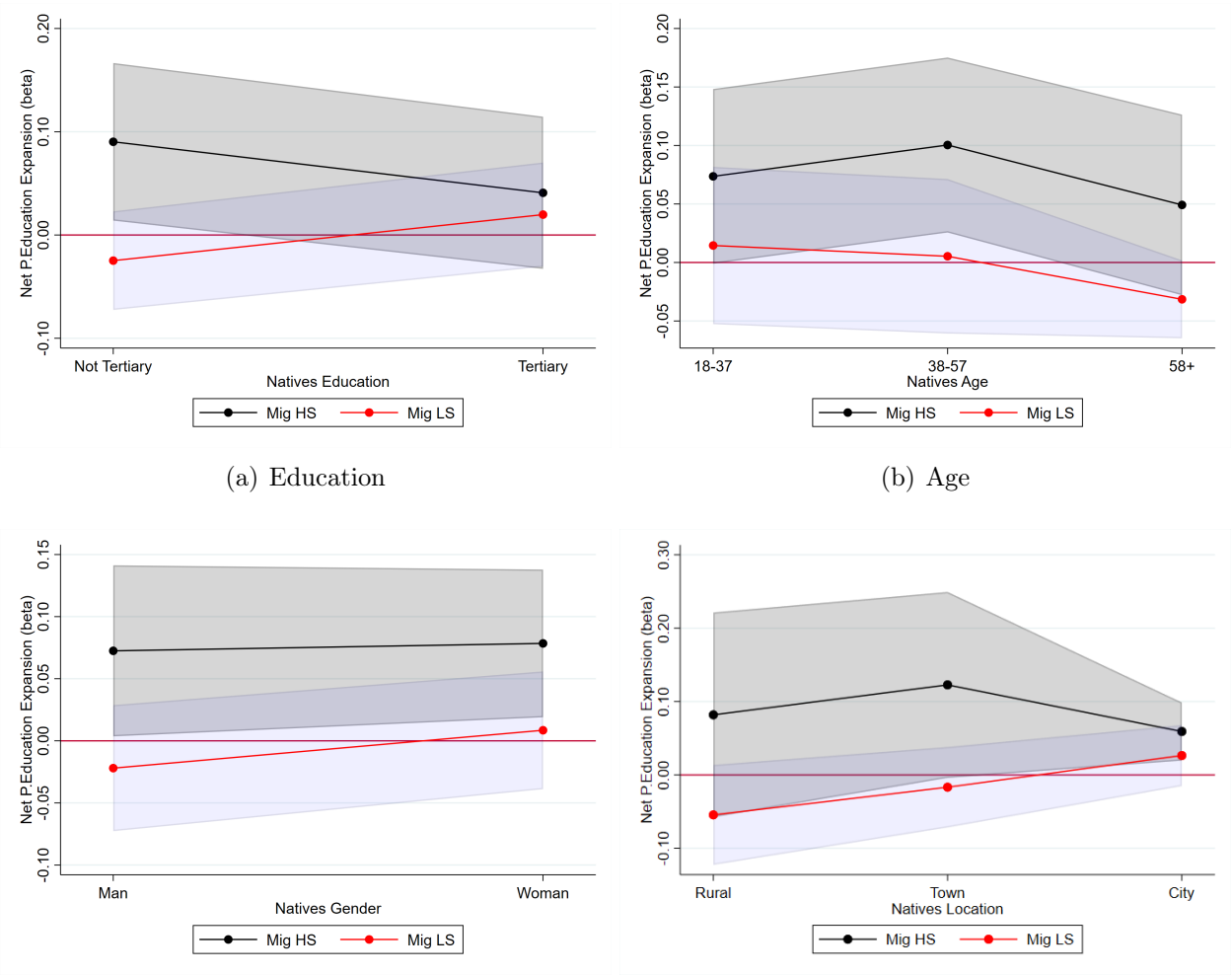
Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. The graphs plot the coefficients of the share of migrants HS and LS estimated on different subsamples by individual characteristics: education (a), age (b), gender (c) and location (d). All the coefficients are estimated with IV estimations. The shadowed area represent the 95% interval of confidence. All the regressions include individual and regional controls, NUTS2 and Election fixed effects. The dependent variable is our standardized measure of net welfare state expansion. The F-stat of the first stage is always above 10. Detailed results are available in Table A4 in the appendix.

Panel (a) of Figure 4.4 reports the results for the measure of net welfare expansion by education level. Interestingly, highly educated natives–on the right side of the chart–respond positively to both high- and less-educated immigration, implying an increased vote share for pro-welfare state parties. The coefficients are precisely estimated. Less-educated voters–

whose coefficients are shown in the left side of the chart–respond to skilled immigration by increasing their vote for welfare state expansion, while their response to less-skilled immigrants is smaller (albeit still in favor of more redistribution). Highly-skilled immigrants are more likely to be considered net contributors to the welfare system, so that more redistribution in response to their inflow is consistent with the interest of natives. Less-educated immigrants are perceived as a net cost to the welfare system, especially by less-educated natives, so one may anticipate a negative response toward the welfare state in elections. However, the estimates are instead small (sometimes not significant) and positive. The age dimension of voters may also be relevant to understanding their vote (Panel (b)). On average, an increase of immigration (either less- or highly-skilled) induces young people (in the age range 18-37) to vote for parties favorable to a welfare expansion. The positive effect of highly-educated immigration decreases voter age increases, becoming not significant for the age 38-57, but the differences are not too large. Panel (c) shows only small differences between male and female responses. Finally, Panel (d) shows that the strongest pro-redistributive effects of highly-educated immigrants on native voters are among individuals in rural areas. The positive response of rural residents to less-skilled immigration is smaller compared to non-rural residents, and non-significant. The fact that the response to high-skilled immigrants pushes voters toward more redistribution while less-skilled immigrants do not have a significant effect on this aspect of the vote is interesting. It appears that individuals move to pro-redistribution parties in response to high-skilled immigrants, but do not significantly alter their preferences in response to less-skilled immigrants. This result is somewhat different from previous findings that less-skilled immigrants drive people to express opinions that are less-favourable to redistribution. We note, however, that in this part of the analysis we are keeping the platform of a party fixed. It is possible that platforms of parties (distinct from votes received) shifted in response to immigration, as we analyze in the next section.

Figure 4.5 has the same structure as Figure 4.4, but with net preference for public education expansion as the dependent variable. Results are rather similar to those in the previous figure. Panel (a) shows that the positive effect of highly-educated immigrants on voting preferences toward public education expansion are stronger for non-tertiary- (less-)educated natives. Immigration does not affect tertiary-educated natives' political preferences on public education much, but less-educated natives respond to high-skilled immigrants by voting for more pro-public school parties. This is consistent with highly-educated natives having stronger preferences for public school spending which are little-affected by increased immigration. Panel (b) shows that the effect of immigration differs only marginally across age groups, and Panel (c) shows almost no difference among by sex. Instead, Panel (d) shows a

Figure 4.5: Net Public Education Expansion and Natives' Individual Characteristics



(a) Education

(b) Age

(c) Gender

(d) Location

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. The graphs plot the coefficients of the share of migrants HS and LS estimated on different subsamples by individual characteristics: education (a), age (b), gender (c) and location (d). All the coefficients are estimated with IV estimations. The shadowed area represent the 95% interval of confidence. All the regressions include individual and regional controls, NUTS2 and Election fixed effects. The dependent variable is our standardized measure of net public education expansion. The F-stat of the first stage is always above 10. Detailed empirical results are available in Table A5 in the appendix.

non-significant anti-public school effect from less-skilled immigrants in rural areas, while immigrants (particularly high-skilled) increase voter preferences for public schooling in urban areas.

Overall, Figures 4.5 and 4.4 display some evidence that the pro-welfare and pro-public school impact of highly-educated immigrants is stronger for rural residents and less-educated voters. These effects, however, are not always statistically significant, and hence should be taken

cautiously and only as suggestive. The small and non-statistically-significant effect of less-educated immigrants on public welfare and schooling is common to most subgroups. While the findings on the effect of highly-educated migrants are consistent with previous results (e.g. Mayda *et al.*, 2018) showing that their inflow may shift opinion towards pro-welfare parties, the zero effects of less-skilled immigrants are somewhat puzzling. In order to shine some light on how less-skilled immigrants may have changed the political preferences for redistribution, we now focus on their impact on party platform.

4.5 Results on Parties' Political Manifesto

Immigration can affect political outcomes not only by influencing voters' preferences, but also by altering parties' political platforms. A large inflow of migrants could force parties to make their positions on immigration (and on other policies affected by immigration, such as redistribution) more salient. They could adjust their manifestos such that parties, rather than voters, are the agents of change in terms of attitudes towards redistribution. Some "pro-redistribution" parties may have evolved over time to favor smaller government (as Section 4.2 showed), and this would amount to a change in policy preferences revealed by votes even if voters did not much change their vote across parties. Our question whether immigration drove some of these changes? To analyze this type of change, Table 4.5 shows the impact of immigration on parties' political preferences. The unit of observation here is a party-year, and the dependent variable is net preference for welfare or public schooling. Panel A presents the estimates related to parties' positions towards net welfare state expansion. Panel B shows the results for preferences toward net public education expansion. As described above, we exploit variation in parties' political platforms between elections over the 2007-2016 period. Our explanatory variables are the less- and high-skill shares of immigrants in the 2005 total population. Each regression, as described in section 4.3.2 includes as controls the logarithm of GDP per capita, the logarithm of the share of tertiary-educated individuals, and the unemployment rate at the country level. They also include country-, party-, and election-year fixed-effects. Standard errors are clustered at the party level.[23] We estimate equation (4.7) using both Least Squares and IV. The F-statistics of the instrument are still strong, but due to the smaller number of observations the estimates are slightly less precise. Columns (1)

[23]The small number of countries in the sample implies that clustering at the country-level, while more conservative, would run the risk of bias due to the small number of clusters. We also run the analysis by clustering at the country-level, and while the estimates and standard errors are similar to those in Table 4.5, the F-statistics of the instrument are significantly smaller. While this suggests caution in interpreting the IV results, the fact that the coefficient and standard error estimates are stable is reassuring. Results are available on request.

and (2) show the coefficient estimates for the overall share of immigrants. Columns (3) and (4) and columns (5) and (6) present the estimates for the shares of high- and less-educated migrants, respectively.

Table 4.5: Political Preferences - Parties

	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
Time	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016
Panel A - Net Welfare State						
Share	-0.029	-0.254*				
	(0.070)	(0.134)				
Share HS			-0.054	0.062		
			(0.307)	(0.494)		
Share LS					-0.036	-0.357*
					(0.088)	(0.183)
Observations	177	177	177	177	177	177
K-P rk Wald F-stat		24.31		21.47		17.39
Adj. R-Square	0.54	0.54	0.54	0.57	0.54	0.54
Panel B - Net Public Education						
Share	-0.101	0.225*				
	(0.070)	(0.134)				
Share HS			-0.196	0.378		
			(0.235)	(0.458)		
Share LS					-0.124	0.261
					(0.094)	(0.181)
Observations	177	177	177	177	177	177
K-P rk Wald F-stat		24.31		21.47		17.39
Adj. R-Square	0.61	0.58	0.61	0.62	0.61	0.59
Party f.e.	✓	✓	✓	✓	✓	✓
Country f.e.	✓	✓	✓	✓	✓	✓
Election f.e.	✓	✓	✓	✓	✓	✓
Country Controls	✓	✓	✓	✓	✓	✓

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Standard errors are clustered at party level. * p<0.1, ** p<0.05, *** p<0.01. As country controls we include the logarithm of GDP per capita, of the share of tertiary educated individuals and the percentage of unemployment individuals. The dependent variables are the standardized net favorable position towards welfare expansion (Panel A) and net favorable position towards public education expansion (Panel B). The analysis is over a sample of parties that are running in multiple elections.

The top panel of Table 4.5 shows that the overall share of immigrants had a significant (at the 10% level) and negative effect on parties' preferences towards welfare expansion. This result suggests that parties' platforms reacted to immigration: a 1 percent increase in the national share of immigrants reduced the average party's preference for net welfare state expansion by 0.25 standard deviations. Looking at the IV estimates of the skill-specific effects of immigration, we find that highly-educated immigration pushed parties toward more

redistribution, consistent with the individual vote results. The coefficients, however, are not precisely estimated, and the effects are statistically insignificant. On the other hand, less-educated migrants push parties away from welfare state support. The IV estimates find a stronger negative and borderline significant impact. The coefficients capturing the effects on parties' preferences towards public education expansion are presented in Panel B of Table 4.5. Overall, we do not find any significant skill-specific effect of immigration on parties' preferences for public schools, as noise seems to prevail.

Table 4.6: Parties Political Preferences - Robustness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Time	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016
Robustness	Above 10% votes				Weighting votes			
Panel A - Net Welfare State								
Share HS	0.261	0.971			0.113	0.023		
	(0.377)	(1.191)			(0.484)	(0.607)		
Share LS			-0.118	-0.475***			-0.156	-0.506**
			(0.125)	(0.177)			(0.112)	(0.213)
Observations	106	106	106	106	176	176	176	176
K-P rk Wald F-stat		2.46		15.94		9.21		13.05
Adj. R-Square	0.35	0.40	0.35	0.37	0.58	0.58	0.58	0.54
Panel B - Net Public Education								
Share HS	0.022	1.664			0.050	0.795		
	(0.351)	(1.291)			(0.421)	(0.719)		
Share LS			-0.022	0.150			-0.108	0.161
			(0.105)	(0.186)			(0.116)	(0.225)
Observations	106	106	106	106	176	176	176	176
K-P rk Wald F-stat		2.46		15.94		9.21		13.05
Adj. R-Square	0.59	0.51	0.59	0.62	0.63	0.61	0.64	0.61
Party f.e.	✓	✓	✓	✓	✓	✓	✓	✓
Country f.e.	✓	✓	✓	✓	✓	✓	✓	✓
Election f.e.	✓	✓	✓	✓	✓	✓	✓	✓
Country Controls	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Standard errors are clustered at party level. * p<0.1, ** p<0.05, *** p<0.01. As country controls we include the logarithm of GDP per capita, of the share of tertiary educated individuals and the percentage of unemployment individuals. The dependent variables are the standardized net favorable position towards welfare expansion (Panel A) and net favorable position towards public education expansion (Panel B). The table show: the analysis over parties with at least 10% votes (col. (1)-(4)) and the analysis using the share of votes as weights (col. (5)-(8)). The analysis is over a sample of parties that are running in multiple elections.

The set of results presented in Table 4.5 shows an overall negative effect of immigration on parties' preferences towards redistribution, driven by the inflow of less-educated immigrants. This is consistent with the findings of Alesina *et al.* (2018). In the party-level regressions of Table 4.5 we gave to each party the same weight. Thus small parties have the same relevance as large parties. We want to account, however, for the larger impact of big parties

and for the fact that small parties may be more prone to change their political agendas. In that case, our simple average over-estimates the changes in political platforms. To limit the effect of small parties on our analysis, columns (1) to (4) of Table 4.6 show the OLS and IV estimates of skill-specific shares of migrants on political preferences over the sample of parties that obtained at least 10% of the votes in at least one electoral event. In columns (5) to (8) we present results when we weight each party by its share of votes. The sample of large parties (above 10%) includes 43 parties, which significantly reduces the size of the sample relative to the full sample of parties which participated at least in two elections (74). Panels (A) and (B) present the results of parties' preferences toward welfare state- and public education expansion, respectively. Our main results are confirmed: an increase in the population share of less-educated immigrants pushes party preferences away from welfare state expansion, and has no significant effect on public education preferences. An opposite but not-significant effect is shown for highly-educated immigrants. The point estimate of the negative effect of less-skilled immigration on preference for redistribution is actually larger in these specifications, implying a 0.5 standard deviation decrease for each percentage point of immigrant increase.

Overall, we take those results as suggestive evidence of an anti-redistribution effect of less-educated immigration on parties' political platforms. During periods of large national inflows of less-skilled immigrants, parties moved away from platforms supporting expansion of public welfare. This is in general true for incumbent parties, as well as for all parties providing the range of electoral choice to voters.

4.6 Conclusion

Will continued growth in immigration lead to the demise of the generous European welfare state? EU citizens may become more reluctant to redistribute, provide public services, and spend on education if they perceive that a large share of the benefits, paid for by taxes, go to immigrants. In this paper, we ask whether the vote of EU citizens, in response to immigration, has evolved in a way that is consistent with this idea. We also qualify this question in two ways: have the EU citizens shifted their vote to anti-redistribution parties over the 2007-2016 period in response to an increased share of immigrants? Or have the parties themselves shifted their agendas? And did the composition of immigrants, between highly- and less-educated, matter in producing electoral response?

Our findings suggest two regularities. First, in contrast with the above prediction, increased highly-skilled immigration pushes the vote of citizens towards parties that have stronger pref-

erences *for* redistribution. Less-skilled immigration, instead, does not significantly change the party-vote of individuals with respect to redistributive preferences. Second, less-skilled immigrants result in the average party moving its agenda toward *less* redistribution.

These two results connect immigration with the content of party vote shares and the agendas of elected parties, and are new in the literature. While other studies have related immigration to opinions revealed by individuals (e.g. Scheve and Slaughter (2001), Alesina *et al.* (2018) and Hangartner *et al.* (2019)), our analysis asks if a change in opinion is reflected by a shift in votes or a shift in a party's programmatic policy, specifically vis-à-vis redistribution. Elected parties are those proposing policies, and their agendas suggest what potential polices will be enacted.

We find these results interesting for two reasons. First, it appears that higher high-skilled immigration may contribute to the preservation of the European welfare state. This is both because highly-skilled immigrants contribute to fiscal balance, and because they motivate political support for the welfare state. An immigration policy that encourages this type of immigration, especially from outside the EU, could be a necessary measure for European society to continue to allow immigration and maintain social cohesion and the welfare state at the same time. Second, we learn that parties' platforms are responsive to the inflow of less-skilled immigrants. As less-skilled immigrants arrive in a country, national parties significantly shift toward platforms that are less-supportive of generous welfare transfers. They seem to aggregate and absorb a preference-shift of voters in response to less-skilled immigration, which is possibly the more visible and controversial form of immigration in Europe. This is consistent with the idea that parties shape their political response to immigration and play important roles in determining the consequences of immigration on European society, and on policies adopted. This paper provides some initial suggestive findings on the connection between immigration and citizens' preferences for welfare, through their votes for elected representatives. We encourage more research in this area, as it is important and crucial for the future of EU policies to understand how citizens respond to immigration with their votes. Even more interesting is the fact that immigration policies that select skills and types of immigrants may indirectly influence future support for the welfare state.

4.7 Appendix

4.7.1 Additional Tables & Figures

Table A1: Correlation between 2000-2007 Growth of Economic Indicators and Post-2006 Predicted Migrants' Growth

	(1)	(2)	(3)
Predicted Mig growth	*All*	*HS*	*LS*
GDP per capita growth rate	-0.09	-0.19*	-0.05
	(0.05)	(0.09)	(0.04)
Observations	96	96	96
Adj. R-Square	0.06	0.16	0.01
Employment growth rate	-0.15	-0.29	-0.08
	(0.11)	(0.18)	(0.10)
Observations	99	99	99
Adj. R-Square	0.04	0.09	0.01
Population growth rate	-0.13	-0.28	-0.08
	(0.11)	(0.18)	(0.09)
Observations	109	109	109
Adj. R-Square	0.02	0.06	0.00
Tertiary growth rate	-0.04	-0.10	-0.02
	(0.03)	(0.06)	(0.03)
Observations	99	99	99
Adj. R-Square	0.02	0.10	0.00

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Standard errors are clustered at country level. * p<0.1, ** p<0.05, *** p<0.01. The table shows the predicted coefficients regressing the growth rate of regional macro indicators between 2000 and 2007 on the regional predicted migration growth with our shift-share strategy over the period 2007 and 2014.

Figure A1: ESS and Actual Voting (Regional Level)

(a) Top 3 parties

(b) Top 5 parties

Note: authors' calculation on ESS and European Election Database (EED). The figure plots the share of votes of top 3 parties (panel (a)) and top 5 parties (panel (b)) in terms of votes in each election available computed with the EED (x-axis) and ESS (y-axis) at regional level. Data are available only for a subset of elections: Austria (2008), Belgium (2010), Denmark (2007), Finland (2007, 2011), Germany (2009), Spain (2008), Sweden (2010), Switzerland (2007, 2011) and United Kingdom (2010). Moreover EDD does not provide the results at regional level for the following regions: Åland (FI20), Helsikini-Ulsima (FI1B) and Northern-Ireland (UKN0).

Table A2: Political Preferences Over 2007-2016 - Individual Voters
Average Preferences Over 2007-2016

	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
Panel A - Net Welfare State								
Share	0.012	0.050***						
	(0.007)	(0.014)						
Share HS			0.030*	0.124***			0.029*	0.118***
			(0.016)	(0.040)			(0.016)	(0.041)
Share LS					0.010	0.051***	0.003	0.014
					(0.010)	(0.018)	(0.010)	(0.017)
Observations	50385	50385	50385	50385	50385	50385	50385	50385
K-P rk Wald F-stat		48.42		33.38		32.72		14.69
Adj. R-Square	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
Share No EU	0.009	0.043***						
	(0.011)	(0.017)						
Share No EU HS			0.023	0.254***			0.019	0.289***
			(0.031)	(0.075)			(0.030)	(0.097)
Share No EU LS					0.009	0.036*	0.006	-0.034
					(0.013)	(0.021)	(0.012)	(0.027)
Observations	50385	50385	50385	50385	50385	50385	50385	50385
K-P rk Wald F-stat		38.80		56.20		37.04		16.17
Adj. R-Square	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
Panel B - Net Public Education								
Share	0.003	0.032						
	(0.010)	(0.020)						
Share HS			0.010	0.135**			0.011	0.147***
			(0.019)	(0.056)			(0.018)	(0.057)
Share LS					0.002	0.016	-0.001	-0.029
					(0.011)	(0.026)	(0.010)	(0.022)
Observations	50385	50385	50385	50385	50385	50385	50385	50385
K-P rk Wald F-stat		48.42		33.38		32.72		14.69
Adj. R-Square	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51
Share No EU	0.003	0.028						
	(0.018)	(0.030)						
Share No EU HS			0.002	0.309**			-0.002	0.401**
			(0.041)	(0.126)			(0.034)	(0.179)
Share No EU LS					0.004	0.008	0.005	-0.090**
					(0.019)	(0.032)	(0.017)	(0.038)
Observations	50385	50385	50385	50385	50385	50385	50385	50385
K-P rk Wald F-stat		38.80		56.20		37.04		16.17
Adj. R-Square	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51
NUTS2 f.e.	✓	✓	✓	✓	✓	✓	✓	✓
Election f.e.	✓	✓	✓	✓	✓	✓	✓	✓
NUTS2 Controls	✓	✓	✓	✓	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Standard errors are clustered at NUTS2 level. * p<0.1, ** p<0.05, *** p<0.01. As individual controls age, gender and education dummies are included. As NUTS2 controls we include the logarithm of GDP per capita, of the share of tertiary educated individuals and the percentage of unemployment individuals. The dependent variables are the standardized net favorable position towards welfare expansion (Panel A) and net favorable position towards public education expansion (Panel B) using the average position of the parties over the time span.

Table A3: Political Preferences over 2007-2016 - Individual Voters
Left to Right Index

	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
Share	0.003	0.013						
	(0.007)	(0.014)						
Share HS			-0.018	-0.031			-0.028**	-0.050*
			(0.011)	(0.029)			(0.012)	(0.030)
Share LS					0.013	0.031	0.019*	0.046*
					(0.010)	(0.020)	(0.011)	(0.024)
Observations	50304	50304	50304	50304	50304	50304	50304	50304
K-P rk Wald F-stat		48.36		32.87		32.58		15.12
Adj. R-Square	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
Share No EU	0.012	0.025						
	(0.011)	(0.019)						
Share No EU HS			-0.004	-0.050			-0.020	-0.124
			(0.026)	(0.073)			(0.025)	(0.076)
Share No EU LS					0.019	0.042*	0.021*	0.072**
					(0.013)	(0.023)	(0.013)	(0.028)
Observations	50304	50304	50304	50304	50304	50304	50304	50304
K-P rk Wald F-stat		38.63		55.30		37.03		16.45
Adj. R-Square	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
NUTS2 f.e.	✓	✓	✓	✓	✓	✓	✓	✓
Election f.e.	✓	✓	✓	✓	✓	✓	✓	✓
NUTS2 Controls	✓	✓	✓	✓	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Standard errors are clustered at NUTS2 level. * p<0.1, ** p<0.05, *** p<0.01. As individual controls age, gender and education dummies are included. As NUTS2 controls we include the logarithm of GDP per capita, of the share of tertiary educated individuals and the percentage of unemployment individuals. The dependent variables is the standardized left to right index using the initial value for each parties over our time span. Higher values implies of the index implies parties collocated on the right side of the political spectrum.

Table A4: Net Welfare State - Individual Heterogeneity Analysis

Estimates	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV
Time	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016
Voters Subsample	*Education*				*Age Groups*					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Not Tertiary	Tertiary	Not Tertiary	Tertiary	18-37	38-57	58+	18-37	38-57	58+
Share HS	0.08**	0.05**			0.09***	0.07*	0.05**			
	(0.04)	(0.02)			(0.03)	(0.04)	(0.02)			
Share LS			0.03*	0.05***				0.06**	0.01	0.04***
			(0.02)	(0.02)				(0.02)	(0.02)	(0.01)
Observations	32579	17725	32579	17725	11126	18399	20779	11126	18399	20779
K-P rk Wald F-stat	15.71	24.81	42.79	43.38	21.25	17.36	17.50	46.95	41.69	42.94
Adj. R-Square	0.71	0.70	0.71	0.70	0.67	0.68	0.74	0.67	0.68	0.74
Voters Subsample	*Gender*				*Location*					
	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
	Male	Female	Male	Female	Rural	Town	City	Rural	Town	City
Share HS	0.08**	0.06***			0.15***	0.03	0.06**			
	(0.04)	(0.02)			(0.05)	(0.04)	(0.02)			
Share LS			0.03*	0.04***				0.02	0.04**	0.05***
			(0.02)	(0.02)				(0.02)	(0.02)	(0.02)
Observations	24850	25454	24850	25454	19794	15678	14832	19794	15678	14832
K-P rk Wald F-stat	19.07	17.97	41.91	45.62	9.44	14.59	17.66	26.28	32.88	54.92
Adj. R-Square	0.70	0.71	0.70	0.71	0.69	0.72	0.70	0.70	0.72	0.70
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
NUTS2 f.e.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
NUTS2 Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Election f.e.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Standard errors are clustered at NUTS2 level. * p<0.1, ** p<0.05, *** p<0.01. As individual controls age, gender and education dummies are included. As NUTS2 controls we include the logarithm of GDP per capita, of the share of tertiary educated individuals and the percentage of unemployment individuals. The dependent variables is the standardized net favorable position towards welfare expansion. The analysis is done on voters characteristics subsamples, based on: education (col. (1)-(4)), age groups (col. (5)-(10)), gender (col. (11)-(14)) and location (col. (15)-(20)).

Table A5: Net Public Education - Individual Heterogeneity Analysis

Estimates	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV
Time	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016
Voters Subsample	*Education*				*Age Groups*					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Not Tertiary	Tertiary	Not Tertiary	Tertiary	18-37	38-57	58+	18-37	38-57	58+
Share HS	0.09**	0.04			0.07*	0.10***	0.05			
	(0.04)	(0.04)			(0.04)	(0.04)	(0.04)			
Share LS			-0.02	0.02				0.01	0.01	-0.03*
			(0.02)	(0.03)				(0.03)	(0.03)	(0.02)
Observations	32579	17725	32579	17725	11126	18399	20779	11126	18399	20779
K-P rk Wald F-stat	15.71	24.81	42.79	43.38	21.25	17.36	17.50	46.95	41.69	42.94
Adj. R-Square	0.56	0.57	0.56	0.57	0.54	0.54	0.60	0.54	0.54	0.60
Voters Subsample	*Gender*				*Location*					
	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
	Male	Female	Male	Female	Rural	Town	City	Rural	Town	City
Share HS	0.07**	0.08***			0.08	0.12*	0.06***			
	(0.04)	(0.03)			(0.07)	(0.06)	(0.02)			
Share LS			-0.02	0.01				-0.05	-0.02	0.03
			(0.03)	(0.02)				(0.03)	(0.03)	(0.02)
Observations	24850	25454	24850	25454	19794	15678	14832	19794	15678	14832
K-P rk Wald F-stat	19.07	17.97	41.91	45.62	9.44	14.59	17.66	26.28	32.88	54.92
Adj. R-Square	0.57	0.56	0.57	0.56	0.57	0.55	0.56	0.57	0.55	0.56
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
NUTS2 f.e.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
NUTS2 Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Election f.e.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on ESS, EULFS, Manifesto Project Database and Eurostat data. Standard errors are clustered at NUTS2 level. * p<0.1, ** p<0.05, *** p<0.01. As individual controls age, gender and education dummies are included. As NUTS2 controls we include the logarithm of GDP per capita, of the share of tertiary educated individuals and the percentage of unemployment individuals. The dependent variables is the standardized net favorable position towards public education expansion. The analysis is done on voters characteristics subsamples, based on: education (col. (1)-(4)), age groups (col. (5)-(10)), gender (col. (11)-(14)) and location (col. (15)-(20)).

Chapter 5

Do emigrants self-select along cultural traits? Evidence from the MENA countries

Abstract[1]

This article empirically investigates whether emigrants from MENA (Middle East and North Africa) countries self-select along two cultural traits: religiosity and gender-egalitarian attitudes. Using Gallup World Poll data on individual opinions and beliefs and migration aspirations, we find that individuals who intend to emigrate to high-income countries exhibit significantly lower levels of religiosity than the rest of the population. They also share more gender-egalitarian views, although this effect holds only among the young (aged 15 to 30), among single women, and in countries with a Sunni minority. For countries most affected by the Arab Spring, the intensity of cultural selection has decreased since 2011. Still, the aggregate effects of cultural selection should not be overestimated. Self-selection along cultural traits has statistically significant but limited effects on the cultural distance between people (i.e., between migrants and natives at destination or between non-migrants in origin and destination countries). Emigration could even reverse the selection effect and lead to cultural convergence if migrants abroad transfer more progressive norms and beliefs to their home country, a mechanism that deserves more attention in future research.

[1]This paper, coauthored with Frédéric Docquier and Ayisit Tansel, has been accepted by International Migration Review

5.1 Introduction

Social scientists have long emphasized that human decisions are affected by norms, values and beliefs found in human societies. The determination and the influence of culture have been abundantly analyzed in cultural sociology, psychology and anthropology (e.g., Alexander (2008); Shweder and Levine (1984); Just and Monaghan (2000)). Recent studies in demography, economics and political science are also converging to rehabilitate culture as a proximate driver of modernization, economic development and democracy (Hammel (1990); Young (2008); Tabellini (2010); Gorodnichenko and Roland (2017); Collier (2017)). Hence, factors that affect the distribution of cultural traits have persistent effects on demographic, economic and political outcomes. International migration is one of these outcomes. Although the existing literature has long emphasized that migrants self-select along educational levels (see, among others, Docquier *et al.* (2007); Grogger and Hanson (2011); Belot and Hatton (2012)), migrants' selection along cultural traits has been understudied.

Focusing on MENA countries, this paper tackles this issue. Using the Gallup World Poll microdata, we construct proxies for religiosity and attitudes towards women's rights, two cultural traits that are correlated with development outcomes and for which MENA countries exhibit distinctive distributions. We then empirically investigate (i) whether these two cultural traits affect the aspirations and plans to emigrate over the 2007-2016 period, (ii) whether selection on culture varies across group of respondents and with country-specific characteristics, and (iii) whether the selection intensity has changed after the Arab Spring. It is worth stressing that our analysis does not make any value judgment about specific cultural traits, and does not argue that cultural differences should be combated or that a set of traits dominates others.[2] Our unique goal is to shed light on the relationship between migration aspirations, preferred destination choices and cultural traits.

There are three important reasons that justify focusing on cultural selection in general, and on selection by religiosity and by gender-egalitarian views in particular. Firstly, cultural selection is one of the main mechanisms through which emigration affects the distribution of cultural traits in the population left behind. If not compensated by ex-post transfers of norms and beliefs from destination to origin countries, selection on gender-egalitarian attitudes is likely to impact effective gender discrimination, which is repulsive in its own rights. In the same vein, the distribution of cultural traits may affect the openness to

[2]Culture shapes the preferences of people, implying that comparisons of demographic or economic outcomes do not reflect comparisons in well-being. This is evidenced in Campante and Yanagizawa-Drott (2015), who show that religious practices in Muslims' countries (as measured by the length of the Ramadan fasting period) have negative implications for economic performance, but increase subjective well-being among the followers.

innovation and the modernization potential of the origin country. Secondly, cultural selection may increase the cultural distance between poor and rich countries, which has been seen as a brake on knowledge diffusion and on the transmission of democratic values. Thirdly, cultural selection is a key determinant of the cultural distance between migrants and host-country citizens, therefore determining the level of cultural diversity at destination, opinions towards immigration, and migrants' capacity to assimilate. The literature on these three potential implications has been growing rapidly for the last two decades or so, as summarized below.

The idea of culture being an important determinant of human, political and economic development probably starts with Weber's theory of the Protestant work ethic (Weber (1946);Guiso *et al.* (2006); Spolaore and Wacziarg (2013)). Culture is seen as a key determinant of individual effort and the overall quality of the institutions that support market friendly exchange. It also plays a role in the formation of trust within and between countries (Guiso *et al.* (2009)). Recent works have produced sound empirical tests of the link between development outcomes and culture, often using opinion survey data as a means of measuring cultural elements such as economic beliefs (Piketty (1995); Di Tella and MacCulloch (2014)), trust (Knack and Keefer (1997); Algan and Cahuc (2010); Felbermayr and Toubal (2010)), etc. The relationship between culture and economic performances has been recently investigated in the Arab context (Kuran (2012); Kostenko *et al.* (2017); Diwan *et al.* (2018)). More related to our work, recent studies focus on the link between views on gender inequalities, religiosity and growth. As far as gender-egalitarian attitudes are concerned, they play a key role in explaining effective discriminations (Bergh (2007); Baxter and Kane (1995)), and the resulting gender inequalities in health, education, political empowerment and employment have long been seen as major barriers to development (UN (2015); Duflo (2012)). As for religiosity, Benabou *et al.* (2015) identify a negative association with individual openness to innovation and with effective patents per capita.[3] Chase (2014) finds a negative association with economic growth, despite the fact that religiosity tends to reduce the level of corruption. Price (2015) shows that MENA countries hold significantly less egalitarian attitudes toward women's employment and specific levels of religiosity compared to those in all other nations sampled. Moreover, the renewed upturn of patriarchal views due to the recent rise of political Islam intensified gender inequality in the MENA context (Tzannatos (1999); Stetter (2008); Alexander and Welzel (2011); Norris and Inglehart (2011); El Mikawy *et al.* (2017)).

Other contributions have highlighted the implications of cultural distance between countries for the speed of knowledge diffusion (e.g. Spolaore and Wacziarg (2012)) and for the democracy transition (e.g. Murtin and Wacziarg (2014)). Spolaore and Wacziarg (2009) measure

[3]It is worth reminding that religion and religiosity are two different concepts. Noland (2005) does not find any negative relation between the share of Muslims in the population and economic performance.

the relatedness between populations using a proxy for cultural distance (based on the probability that two randomly drawn individuals from the two populations share identical genes). They show that genetic distance captures barriers to the diffusion of development, while similarities in genetic traits would tend to facilitate communication and understanding, enhancing the diffusion and adaptation of complex technological and institutional innovations. Desmet *et al.* (2011) and Spolaore and Wacziarg (2015) show that populations that are more closely genetically tend to be more similar with respect to traits (habits, customs, beliefs, values, etc.) and share closer technological and political characteristics.

Finally, migrants' selection on culture determines the level of cultural diversity in the host country. Although diversity induces beneficial effects on the host country (Ottaviano and Peri (2006); Alesina et al. (2016)), several empirical studies show that immigrants' economic and socio-demographic outcomes at destination depend on the distance between their identity and the dominant norms (Pendakur and Pendakur (2005); Fernandez and Fogli (2009); Alesina and Giuliano (2010); Battu and Zenou (2010); Casey and Dustmann (2010); Bisin *et al.* (2011); Islam and Raschky (2013)). The effect is usually negative and its size is uncertain. Perceived cultural distance is also the source of negative attitudes towards immigrants (Card *et al.* (2006)), leading to discrimination, marginalization and exclusion from the economic, social and political life. In particular, Islamophobia has been increasing in Western societies, and around 70% of western natives think that tensions between the Muslim and Western worlds originate in cultural and religious differences (Gallup (2010)).

In sum, migrant selection on culture has potentially important effects on many outcomes. Many papers have emphasized the effect of the average cultural distance between countries on dyadic migration flows (Mayda (2006); Belot and Ederveen (2012); Krieger *et al.* (2018)). However, to the best of our knowledge, there are very few papers investigating the link between individual cultural traits and migration aspirations using micro-data. Berlinschi and Harutyunyan (2018) use data from the Life in Transition Survey (LITS) jointly collected by the European Bank of Reconstruction and Development and the World Bank in 2010. It covers around 39,000 households from 34 countries. They identify a positive correlation between migration aspirations from Eastern European or post-Soviet countries and opinions about home-country governance, political participation and trust in other people. More related to our analysis, Myers (2000) finds that migration aspirations of US citizens are negatively correlated with involvement in social activities related to religion. Hoffman *et al.* (2015) finds that external religiosity (e.g., participating in religious activities) and internal religiosity (e.g., spirituality) induce different effects on migration aspirations of Roman Catholic Mexican students. Using the second (2010-11) and third (2012-14) waves of the

Arab Barometer for nine Islamic countries, Falco and Rotondi (2016) focus on the role of radical Islam views measured by opinions about the use of Islamic law in the formulation of penal, personal status and inheritance law. They find that radical views are negatively correlated with migration aspirations. Our paper uses a similar specification but focuses on different origin countries and on specific cultural traits that are often considered as causing cultural tensions in Western societies. Taking advantage of the coverage of the database, it sheds light on the heterogeneity – across groups of individuals and across countries – of cultural selection patterns in the MENA region and on the role of internal conflict.

The rest of this paper is organized as following. Section 5.2 explains our research objectives and describes the data sources used to proxy migration aspirations and culture. The empirical specification and econometric issues are discussed in Section 5.3. Estimation results are presented in Section 5.4. The social and policy implications of our results are discussed in Section 5.5. Finally, Section 5.6 concludes.

5.2 From theory to data

There is a large literature in sociology and demography investigating the determinants of country- or group-specific aspirations to migrate. To mention just a few, Becerra *et al.* (2010) and Becerra (2012) show that lower socioeconomic status is not enough to explain intentions to emigrate of Mexican adolescents in Tijuana; other factors like perceived discrimination and pre-migration acculturation also play a key role. In a sample of 17 Latin American countries, Wood *et al.* (2010) show that family intention to migrate to the US increased after being victim of a crime. De Jong *et al.* (1996) and Drinkwater and Ingram (2009) show that personal values are relevant to explain internal willingness to migrate of Thais and Britons. These results highlight the importance of noneconomic factors as determinants of migration intentions.

In line with the above studies, we analyze individual emigration as a two-step process. The first step relates to migration aspirations: "aspiring" migrants express an intention/desire to emigrate and search for migration opportunities, while the others decide to stay put. The second one relates to the realization of these aspirations: some searchers in the pool of aspiring migrants find opportunities to migrate (such as a job offer, a visa, etc.). Our model focuses only on the first step of the process. Theoretically speaking, we model migration aspirations of each individual i living in region r at year t as depending positively on the returns and negatively on the costs of migrating (as in Chort (2014) and Ruyssen and

Salomone (2018)). It also depends on the availability of information about economic and social conditions abroad. Returns, costs and information sets vary across individuals. They are affected by a set of individual observed characteristics (X_{irt}), and by a random component (ϵ_{irt}) that captures unobserved heterogeneity in preferences. The net benefit from migrating, NB_{irt}, can be written as:

$$NB_{irt} = f(X_{irt}) + \epsilon_{irt}, \tag{5.1}$$

where $f(.)$ is a general function that determines the net returns from migrating.

Hence, the decision to become an aspiring migrant can be expressed as:

$$m_{irt} = \begin{cases} 1 \;\; if \;\; NB_{irt} > 0 \\ 0 \;\; if \;\; NB_{irt} \leq 0 \end{cases} \tag{5.2}$$

In the empirical section, we will assume that ϵ_{irt} follows a logistic distribution and that $f(X_{irt})$ is a linear function of individual characteristics. Under these assumptions, the probability that individual i becomes an aspiring migrant is governed by a Logit model:

$$\Pr(m_{irt} = 1) = \Phi(X_{irt}\beta) \tag{5.3}$$

where $\Phi(.)$ is the logistic function and β is a set of parameters to be estimated.

The vector X_{irt} includes standard regressors (e.g., education, age, gender, presence of family members or friends abroad, etc.).The literature on the drivers of aggregate migration stocks and flows has identified an effect of cultural distance between countries (Krieger *et al.* (2018)), which can be correlated with many unobserved factors. At the micro level, we hypothesize here that cultural traits influence the returns and costs from migrating. The distance between individual cultural traits and the host-country culture can affect the net returns from migrating through labor market discrimination abroad, racism and violence, incapacity to worship, adjustment to host-country norms, etc. (see Adsera and Pytlikova (2015); Falck *et al.* (2012); Falck *et al.* (2018))

To test this hypothesis, we need micro data on migration aspirations, cultural traits and other individual characteristics. Such data can be obtained from the representative Gallup World Polls (GWP). Although the GWP database covers 148 countries, our sample is limited to 17 MENA countries where Gallup conducted at least one wave of its survey between the

years 2007 and 2016.[4] On average, the sample includes about 1,000 randomly selected respondents per year and per country. For the majority of countries in our sample, the data are collected through face-to-face interviews. Exceptions are Iran and Iraq, where interviews were mainly conducted through phone calls.[5] The sampling frame is such that GWP data are representative of the entire population aged 15 and over (including populations from rural areas). Our full sample includes 60,284 native respondents on the working age population only (i.e., individuals aged 15 to 64).[6]

Measuring migration aspirations. – The GWP includes several questions capturing migration aspirations, preferred destination choices, and whether individuals are actively taking steps to emigrate. The fact that the GWP database covers many countries makes it exceptional. As the data are relatively new, the literature relying on these data to capture migration aspirations is limited. Manchin *et al.* (2014) investigate the impact of individual satisfaction with local and country-level amenities on the willingness to migrate internationally and locally. Dustmann and Okatenko (2014) study the role of wealth constraints and local amenities in governing migration intentions from sub-Saharan African countries. Docquier *et al.* (2014) and Dao *et al.* (2018) study the determinants of migration aspirations after aggregating GWP data by country pair and by education level. Docquier *et al.* (2015) use the GWP data to proxy the number of potential migrants who could respond to an abolition of migration barriers. Bertoli and Ruyssen (2018) quantify the effect of migrant networks on the migration aspirations and destination choices. Ruyssen and Salomone (2018) investigates whether gender discrimination fosters women's migration aspirations.

The three GWP questions used in this paper are the following:[7]

- Ideally, if you had the opportunity, would you like to move permanently to another country, or would you prefer to continue living in this country?
- To which country would you like to move?
- Are you planning to move permanently to another country in the next 12 months, or not?

[4]Our sample of MENA countries includes Afghanistan, Algeria, Azerbaijan, Chad, Egypt, Jordan, Iran, Iraq, Lebanon, Mali, Mauritania, Morocco, Niger, Palestine, Syria, Tunisia, and Yemen. It excludes the Persian Gulf countries, which exhibit much greater levels of income and lower shares of aspiring migrants (3.5% compared to the sample mean around 24%). Nevertheless, including Kuwait and UAE (for which GWP data are available) in our sample does not affect the results, as shown in Table A10 in the Appendix.

[5]In these two countries, more than 80% of the population have a telephone land-line.

[6]Descriptive statistics and correlation coefficients are provided in the 5.7.3. See Tables A5 and A7.

[7]In the GWP, these questions are coded *wp1325*, *wp3120* and *wp10252*, respectively.

We define as aspiring migrants those who answer the first question affirmatively. Note that the last two questions are asked to aspiring migrants only. In line with Bertoli and Ruyssen (2018), we find that migration aspirations are correlated with actual migration flows. Using the annual flow data from the OECD *International Migration Database*, the correlation with aspirations is positive (0.435) and significant at the 1% threshold. Hence, patterns of migration aspirations are likely to be similar to the patterns of actual migration.

The average share of aspiring migrants in our sample is around 24%. Large variations exist across countries. Countries exhibiting the greatest shares of aspiring migrants are Syria (35.6%), Jordan (27.6%) and Algeria (27.5%); countries exhibiting the smallest shares are Niger (16.1%), Azerbaijan (18.3%) and Chad (17.9%). Cross-country variations in destination choices are even larger. Most MENA countries were colonized or administered by European powers. Algeria, Chad, Lebanon, Mali, Mauritania, Morocco, Niger, Syria and Tunisia were colonized or administered by France during the 19th and 20th centuries. Egypt, Iraq, Jordan and Palestine and (South-)Yemen were colonized or administered by the Great Britain. Azerbaijan proclaimed independence from the Soviet Union in 1991. Due to linguistic proximity and network effects, these historical ties are still affecting the preferred destination of aspiring migrants. On average, the share of aspiring migrants who would like to emigrate to an OECD destination country equals 52.3%.[8] The latter share amounts to 90.9% in Morocco and to 86.7% in Algeria; it only equals 10% in Yemen and 12.8% in Niger.

Measuring cultural traits. – Our objective is to test whether cultural traits affect the aspiration to migrate. The GWP database includes several questions on cultural norms, beliefs, values and attitudes. Being aware that culture is a broad and vague concept (Shenkar (2012)), we proxy cultural traits using five questions which cover attitudes towards religiosity and gender-egalitarian views in the MENA region:

$Q1$ Is religion an important part of your daily life?

$Q2$ Have you attended a place of worship or religious service within the past seven days?

$Q3$ Do you agree that women and men should have equal legal rights?

$Q4$ Do you agree that women should be allowed to hold any job for which they are qualified outside the home?

$Q5$ Do you agree that women should have the right to initiate a divorce?

[8] Tables A1 and A2 in on line 5.7.1 lists the top-5 desired destinations by country of origin, distinguishing between OECD and non-OECD destinations.

The first two questions are asked in all countries of the world, while the other three are asked in specific geographical regions. We normalize responses between 0 and 1 such that sharing gender-egalitarian views and not being religious are coded as 1.

Several methods can be used to extract synthetic information on cultural traits.[9] We conduct a *Principal Component Analysis* (referred to as PCA henceforth). The PCA data reduction method allows identifying linear combinations of questions that explain the greatest share of variance in cultural traits. The results of the PCA are provided in 5.7.2. The first two components of the PCA explain the highest amount of variance. The first one reflects gender-egalitarian views whereas the second one reflects religiosity. For each of these two traits, we combine individual responses to the GWP questions using the values of the eigenvectors as weights.[10] This defines our benchmark indices of *Gender-egalitarian attitudes* and *Religiosity*.

Other methods can be used to aggregate cultural questions. In the Appendix, we consider four alternative methods and obtain very similar results (see Table A11). Firstly, Factor Analysis (FA) identifies latent variables that explains common variation underlying observed data, rather than explaining an optimal amount of variance (see Preacher and MacCallum (2003)). We run a FA on our data using a principal factor method, since the normality of data is not granted (see Costello and Osborne (2005)). Secondly, we run a Multiple Correspondence Analysis (MCA), which is similar to PCA but specifically adapted in the case of categorical variables (see Tuccio *et al.* (2016)). Thirdly, we remove the linearity assumption of the PCA and perform a Kernel Principal Component Analysis (KPCA), allowing for non-linear data reduction. Fourthly, we assign the same weight to each question and compute the average response (AVG). Each method generates new indices of gender-egalitarian attitudes and religiosity. Differences across indices are discussed in Table A5. Tables A8 and A9 show that the correlation between the new indices and PCA ones is always large and highly significant. Given the strong robustness of our findings, all results presented in the core of the paper rely on the PCA aggregation method.

Table 5.1 reports the mean value of each indicator by country. Lebanon and Azerbaijan are the most progressive in terms of gender-egalitarian attitudes. This may be due to the fact that Azerbaijan has a Soviet legacy and Lebanon has about 44% of Christians, making it a pluri-confessional society with 17 recognized religious denominations and cultural groups

[9]For example, Spolaore and Wacziarg (2015) combine different sets of questions in order to create an aggregate measure of culture. They use the Euclidean distance to aggregate question-specific differences between groups of respondents.

[10]A similar technique is used in UN (2005). Pairwise correlations between components and responses to the GWP questions are provided in Table A6.

Table 5.1: Cultural traits: mean levels by country

Countries	Gender	Religiosity	Countries	Gender	Religiosity
Afghanistan	0.502	0.156	Mauritania	0.684	0.146
Algeria	0.732	0.232	Morocco	0.675	0.234
Azerbaijan	0.846	0.633	Niger	0.552	0.104
Chad	0.767	0.098	Palestine	0.686	0.272
Egypt	0.738	0.188	Syria	0.668	0.245
Iran	0.8	0.342	Tunisia	0.739	0.355
Iraq	0.594	0.311	UAE	0.805	0.163
Jordan	n.a.	0.691	Yemen	0.676	0.237
Kuwait	0.761	0.137	All	0.699	0.256
Lebanon	0.853	0.313	Main Insurgents	0.713	0.267
Mali	0.644	0.164	Others	0.691	0.252

Source: Authors' calculation on the Gallup World Poll. All the values in the table are the mean values of each indicator. Main Insurgents: Algeria, Egypt, Syria, Tunisia and Yemen.

(Karouby (2014)).[11] Iran, Tunisia and Azerbaijan are the less religious countries.[12] Finally, sub-Saharan African countries (Chad, Mauritania, Mali and Niger) exhibit a high level of religiosity. The geographical distribution of these cultural traits is plotted in Figures 5.1 and 5.2.

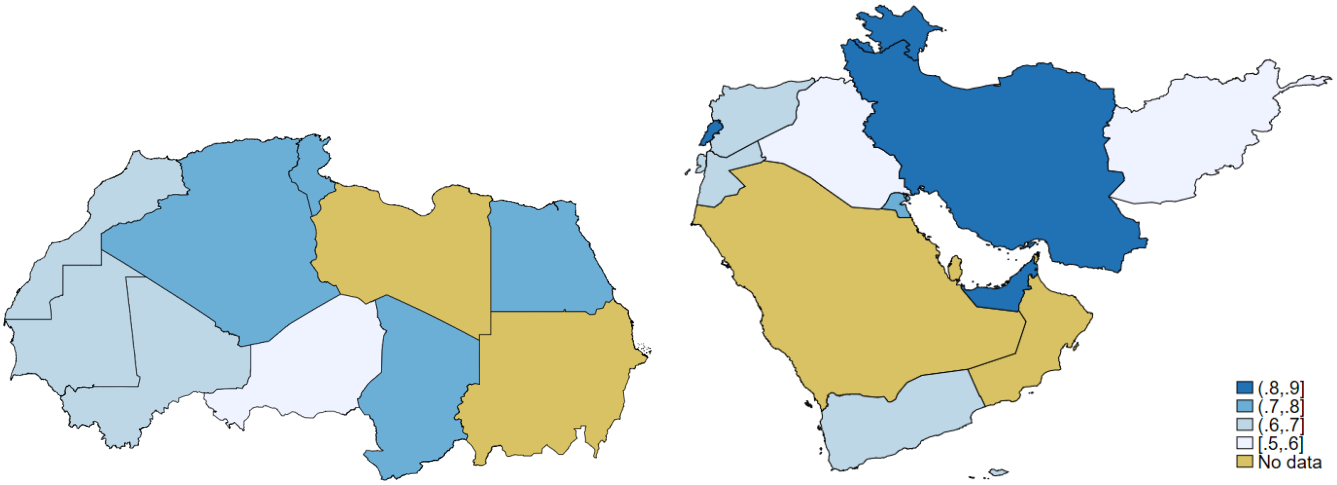
GWP is not the only database documenting the distribution of cultural traits. For example, several questions of the World Values Survey (WVS) can be used to document beliefs and values. However, the WVS includes a smaller set of countries,[13] and has no specific question on migration plans and aspirations. Still, some WVS questions closely relate to our two indices of cultural traits. As far as religiosity is concerned, the sixth wave of the WVS includes four questions: (i) How important in life is religion; (ii) How often do you attend religious service?; (iii) How often do you pray?; (iv) Are you a religious person? As for gender-egalitarian attitudes, the WVS includes two questions on gender-egalitarian views: namely (i) When jobs are scarce, men should have more right to a job than women?; (ii) On

[11] It is worth noticing that countries with gender-egalitarian institutions like Tunisia and Morocco (see Kammoun (2014) and Sadiqi (2014)) are not the most gender-egalitarian countries.

[12] Moaddel and Azadarmaki (2002) indicate that Iranians are not more religious than other Middle Eastern populations. Kashavarzian (2010) concludes that average levels of religiosity have remained constant compared to the pre-1979 revolution period, and that participation in Friday prayers has declined. In Azerbaijan, the major political forces are secular. The country has a high level of literacy and human development, a legacy of the Soviet period.

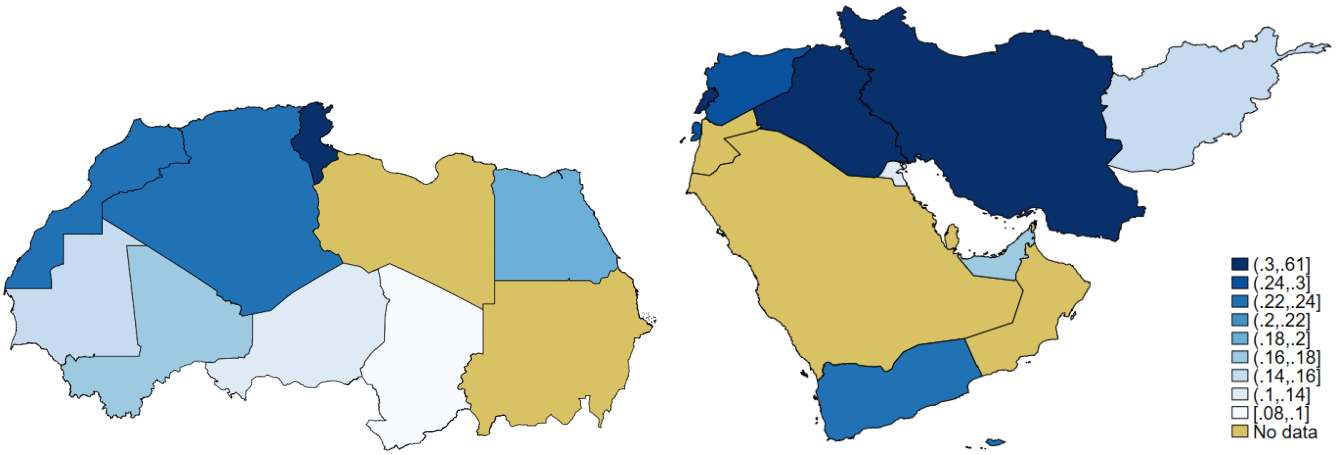
[13] As far as the MENA region is concerned, WVS data are available for the following countries: Algeria, Azerbaijan, Egypt, Jordan, Iraq, Lebanon, Morocco, West Bank and Gaza, Tunisia and Yemen.

Figure 5.1: Cross-country heterogeneity in gender-egalitarian views



the whole, men make better political leaders than women do? We normalize WVS responses between 0 and 1, using the same order as before.[14] The upper panel of Table 5.2 reports the correlations between the country-specific mean levels of our indicators and of the WVS data. Our index of religiosity is highly correlated with the WVS responses. Our gender-egalitarian index is poorly correlated with the WVS index of economic equality. It is however well correlated with the WVS index of equality in politics.

Figure 5.2: Cross-country heterogeneity in religiosity



Correlates of cultural traits. – We finally investigate whether our proxies for cultural traits

[14]Being religious takes value of 0, while not being religious value 1; sharing gender-egalitarian views is associated with value 1 and 0 otherwise.

Table 5.2: Correlation between aggregate cultural GWP, WVS and macro indices

Indicator	Gender		Religiosity	
WVS Indices				
Important in life: religion	-	-	-	0.832***
Attending religious service	-	-	-	0.843***
Praying frequently	-	-	-	0.718**
Religious Person	-	-	-	0.947***
Gender equality: job	-	0.195	-	-
Gender equality: politics	-	0.646**	-	-
Macro Indicators				
Sunnis	-0.296	-	-0.456*	-
Shiites	0.366	-	0.651***	-
ln Gdp	0.644***	-	0.693***	-
Rule of Law	0.259	-	0.101	-
Control Corr.	0.157	-	-0.049	-
Network	0.345	-	0.199	-

Source: Authors' calculations based on the Gallup data, CIA World Factbook, World Bank indicators, Maddison Project, UN databases and World Value Survey. For the World Value Survey data the list of countries available is the following: Algeria, Azerbaijan, Egypt, Jordan, Iraq, Lebanon, Morocco, Palestine, Tunisia and Yemen. *** p<0.01, ** p<0.05, * p<0.1.

are correlated with six macro indicators capturing the branch of Islam (Sunni or Shia) that is prevalent in the origin country, the level of economic development, the quality of institutions, and past migration flows. Data on the shares of Sunnis and Shiites in the Muslim population are taken from the CIA World Factbook and from the PEW Research Center; data on GDP per capita are obtained from the Maddison Project; data on the control of corruption and on the rule of law are taken from the Worldwide Governance Indicators of the World Bank; as for past migration, we compute the percentage of respondents with a family member or a friend abroad from the GWP data (a proxy for migration networks).

Correlation coefficients are reported in the lower panel of Table 5.2. Two main findings emerge from this table. First, our index of religiosity is highly correlated with the composition of the Muslim population. Countries with a greater share of Shiites are less religious than Sunni countries. Second, the level of development is highly correlated with religiosity

and gender-egalitarian attitudes, in line with the empirical literature on culture and economic growth (see Duflo (2012); Benabou *et al.* (2015); Chase (2014)). The strong relationship between religiosity, gender-egalitarian attitudes and economic development is the main reason why we focus on these two cultural traits. On the contrary, our indices are poorly correlated with the quality of institutions and with previous existing emigrant stocks.

5.3 Empirical Strategy

Our goal is to analyze the determinants of migration aspirations, and to test whether these aspirations are affected by cultural traits. This section describes the benchmark empirical specification and discusses some econometric issues.

Benchmark specification. – As explained in Section 5.2, our benchmark empirical model features the intention to migrate as the dependent variable. For respondent i originating from region r at year t, the variable m_{irt}, takes a value of 1 if individual i expresses a desire to migrate abroad permanently and 0 otherwise. In line with Eq. (5.3), we opt for a logit model:

$$Prob\{m_{irt} = 1|Culture_{irt}, X_{irt}\} = \Phi\{\alpha + \beta Culture_{irt} + \Gamma X_{irt}\} \tag{5.4}$$

where $Culture_{irt}$ is our proxy of cultural traits ranging from zero (when they do not share gender-egalitarian views or when they are religious) to one; X_{irt} is the set of other determinants of migration aspirations that vary across individuals; α , β and Γ are the parameters to be estimated. We cluster the error terms at the country level.

The set of control variables includes: age, gender, marital status, the presence of children in the household, the level of income per household member and its squared, the education level (a dummy variable equal to one if the respondent has least 9 years of education), and the presence of a friend or relative abroad. These variables are denoted by x_{irt}, a subset of X_{irt} (as explained below). In line with the existing literature, these variables affect the size of migration costs as well as the expected gains from migration.

We explain above that migration aspirations are correlated with actual migration flows, in line with Bertoli and Ruyssen (2018). However, the number of aspiring migrants is much greater than the number of actual migrants. Hence, cultural traits and other determinants may affect realization rates, and may have heterogeneous effects on the desire to emigrate and on the capacity to realize these aspirations. Hence, as a robustness check, we also estimate

Eq. (5.4) using migration plans (instead of migration aspirations) as a dependent variable. This dependent variable takes a value of 1 if respondent i is actually making steps to move to another country within 12 months, and 0 otherwise.[15]

Eq. (5.4) is estimated using the sample of working age respondents living in the 17 MENA countries, and using the main cultural proxies identified in Section 5.2 (i.e., the synthetic indices of gender-egalitarian attitudes and religiosity).[16] In addition, our estimates can be affected by the presence of immigrants in the sample; the latter are likely to exhibit different characteristics, cultural traits and have different migration strategies (e.g., transit immigrants). To keep the sample as homogenous as possible, we exclude the foreign born from the sample and only consider native residents.

Our approach entails several methodological issues that might lead the logit model to generate inconsistent estimates. In particular, we discuss below how we deal with heterogeneous effects and with endogeneity problems.

Heterogeneous effects. – Migration patterns in general, and the role of cultural traits in particular, may vary according to the regional context, to the choice of destination, or to individual characteristics. To test for heterogeneity across countries of origin, we first estimate Eq. (5.4) at the country level. Since we find large variations across countries, we augment Eq. (5.4) with some country-specific variables and with their interaction with cultural proxies. In line with Table 5.2, the set of country-specific characteristics includes the country shares of Sunnis and Shiites among the Muslim population, for the log of GDP per capita, for two indicators of institutional quality, and for the share of native citizens living in a OECD country member state.

Secondly, to test for heterogeneity across periods, we distinguish between the pre-Arab Spring period and the subsequent years. The Arab Spring started in December 2010 in Tunisia (with the attempted self-immolation of Mohamed Bouazizi); it triggered riots and political unrest in several MENA countries in the subsequent months. Most economies were adversely affected in the post-Arab Spring period. Investments collapsed, tourism and exports declined, capital flight accelerated (especially in countries such as Egypt, Tunisia and Yemen). In Syria, economic production declined significantly. Unemployment increased throughout the region and economic growth was negative in Tunisia and Yemen (Richards *et al.* (2014)). The

[15]The question writes as: *Are you planning to move permanently to another country in the next 12 month, or not?* It is only asked if the answer related to the intention to migrate is affirmative.

[16]Remember we use PCA to construct our indices of cultural traits. As explained in Section 5.2, using alternative methods gives similar results. The results of these variants are presented in Table A11 in Appendix, where we only consider migration aspirations to OECD destination countries.

Arab Spring may have impacted cultural norms and migration intentions jointly. Gill (1999) notes that socio-econonomic crisis may cause a loss in cultural identity, making individuals more susceptible to religion, as we observe in our case. In particular, the political instability and the rise of authoritarianism that characterize the post-Arab Spring period (sometimes referred to as the Arab Winter) may have affected the process of cultural selection of aspiring migrants.[17] We exploit this possible source of variation by focusing on the first GWP question on religiosity (see $Q1$ above), which is the only question that spans all years from 2007 to 2016. The correlation between responses to $Q1$ and our synthetic indicator of religiosity is large (0.688) and highly significant.[18] We then augment Eq. (5.4), estimating the following equation:

$$\begin{aligned} Prob\{m_{irt} = 1 | Relig_{irt}, X_{irt}, AS_{rt}\} = \Phi\{\alpha + \beta Relig_{irt} + \delta Relig_{irt}.AS_{rt} \\ + \zeta AS_{rt} + \eta AS_{rt}.x_{irt} + \Gamma X_{irt}\} \end{aligned} \tag{5.5}$$

where AS_{rt} is a dummy variable that captures the post-Arab Spring period (i.e., a dummy equal to one for the years 2011 to 2016) and $Relig_{irt}$ is a dummy variable that takes value of 1 if religion is an important part of daily life for individual i from region r at year $t \in \{2007, 2008, .., 2016\}$.

The estimated coefficients β and δ capture the correlation between religiosity and intention to migrate, and the effect of the Arab Spring on such correlation, respectively. If $\hat{\delta}$ is positive and significant, this implies that intending migrants are less religious after the Arab Spring.[19] We are fully aware that the Arab Spring phenomenon was extremely heterogeneous across countries. Governments were overthrown in some countries (e.g. Tunisia) while the Arab Spring had smaller effects in other countries (e.g. Mauritania). To account for this heterogeneity, we estimate Equation (5.5) on the full set of countries, on a sample of countries that were highly impacted by the Arab Spring (i.e. *Main Insurgents*[20]) and on

[17] In the post-Arab Spring period, Islamists became dominant in Egypt and Tunisia, and partly in Yemen. Richards *et al.* (2014) see two reasons for this. First, the fear of political Islam declined. Second, Islamist parties propagated on the lack of social justice and corruption under the previous rulers. Initially, their messages became appealing to the middle class voters. Later, they proliferated due to grassroot movements.

[18] The correlation between our aggregate indices and the responses to GWP questions are provided in Table A6

[19] Although apparently similar to a difference-in-difference strategy, we could hardly define our estimation as a standard DiD because the definition of the groups (being religious or not) is not related to the treatment (after Arab Spring), while in the standard DiD strategy treatment and control group are defined on the treatment itself.

[20] It includes Algeria, Egypt, Tunisia, Syria and Yemen. We better explain the selection of those countries as Main Insurgents in Section 5.4.2

the other MENA countries. We would expect a stronger and more significant effect of Arab Spring among the *Main Insurgents* countries. Moreover, those countries may exhibit specific distributions of cultural traits and migration aspirations.

Thirdly, at the individual level, we investigate whether cultural selection is affected by the intended destination of aspiring migrants. If migrants have cultural values that are more similar to those of the intended host country, we expect to find heterogeneous selection patterns across preferred destination types. To deal with this issue, we estimate our model with a modified dependent variable, distinguish between migration aspirations towards the OECD member states or towards non-OECD destinations. Furthermore, we perform several robustness checks by sub-samples, distinguishing between age groups (15-30, 31-45, 45-65), between skill groups (respondents with less than 9 years of education or more), between employment status (unemployed, employed, out of the labor force), between gender and marital status (married and unmarried individuals, female and male) between religious groups (Muslims, Christians, others) and by place of residence (farm, town and city).

Omitted variables. – Although we control for a traditional set of individual characteristics, migration aspirations can be governed by unobserved characteristics. These omitted variables can be related to the regional environment of the respondents (e.g., governance and security in the region, ethnic composition of the population, climatic conditions, distribution of cultural traits, percentage of native citizens abroad, etc.), or to their own characteristics (e.g., cognitive skills and abilities, family ties, etc.). These unobserved characteristics may jointly affect the acquisition of cultural traits and migration aspirations (see Bisin and Verdier (1998)). To deal with unobserved regional characteristics, we take advantage of the fact that GWP identifies the detailed geographical location of the respondent (within the country), and covers several years. For these reasons, we systematically augment the set of controls with spatial and year (or GWP wave) fixed-effects, and estimate a fixed-effect logit model. Hence, the full set of control variables in Eq. (5.4) writes as:

$$\Gamma X_{irt} = \theta_r + \theta_t + \gamma x_{irt},$$

where θ_r and θ_t are the intra-country region and year/wave fixed effects,[21] and γ is the vector of incidence parameters related to the observed controls.

As for unobserved individual characteristics, our specification could still lack unobserved factors that could jointly influence migration aspirations and cultural traits. Our estimated

[21] We estimate our model with fixed effects estimates, since the results of the Hausman test reject the random effects key assumption.

coefficients would then capture spurious correlations. Moreover, disparities in the distribution of covariates between aspiring migrants and non-migrants may influence the accuracy of our estimates. As shown by Imbens and Rubin (2012), large covariates distributional gaps magnify the sensitivity of the estimated coefficients to any ostensibly minor change in the specification. To address this issue, we use a matching technique. We implement a design phase that precedes the empirical analysis, and which consists in constructing a balanced sample in terms of observed covariates. In practice, we match aspiring migrants with non-migrants using the *Mahalanobis Metric Matching* method, using all observed covariates to compute the distance between individuals. A particular property of the *Mahalanobis Metric Matching* method is that the resulting set of matches is invariant to affine transformations of the covariates. Moreover, it generates samples such that the number of aspiring migrants and non-migrants are equalized in each country. This method allows us to compare intending and nonintending migrants that are really alike in the whole set of observables, apart from their cultural traits. If unobserved factors are correlated with observed factors, we can also assume that a reduction of observable differences will also reduce unobservables differences, minimizing this potential source of bias (see Altonji *et al.* (2005) and Oster (2019)). Moreover, if cultural traits are exogenous,[22] we can claim a certain degree of causality to our estimated coefficients. Otherwise, the *Mahalanobis Metric Matching* method simply identifies a cleaner correlation. Since unobserved factors can be also relevant in the estimation of Eq. (5.5), we use the same method to build a balanced sample in terms of covariates, matching individuals between pre- and post-Arab Spring phenomenon. We then conduct our regressions on the balanced samples, making the estimates more credible and robust.

5.4 Results

Our empirical analysis follows the structure explained in Section 5.3. In Section 5.4.1, we begin by investigating the effect of cultural traits on migration aspirations using fixed-effects logit regressions, and distinguishing between OECD and non-OECD destinations. In addition, we check whether similar cultural selection patterns can be identified when considering short-run migration plans (instead of migration aspirations). In Section 5.4.2, we perform several robustness checks to explore how our results are influenced by individual characteristics (see subsection 5.4.2), and whether they are robust to the matching technique (see subsection 5.4.2). We account for heterogeneous effects across countries in subsection

[22] In their survey of the literature, Guiso *et al.* (2006) identify cultural traits, in particular religiosity, as those who remain fairly unchanged over time.

5.4.2. Finally, subsection 5.4.2 investigates the effect of Arab Spring on cultural self-selection.

5.4.1 Logit Regressions

Table 5.3 focuses on migration aspirations and describes the results of the fixed-effect logit regressions for the full sample of MENA countries and by type of destination. Columns (1) and (2) report estimates for migration aspirations to all destinations. It shows that aspirations are correlated with religiosity. The coefficient of religiosity is positive (0.389) and significant at the 1% threshold. The logit model is non linear. To illustrate the magnitude of this effect, let us define the benchmark category of respondent as males with college education, aged 24 to 35, married with children, without friends or relatives abroad, and with a level of religiosity equal to the sample mean (0.259). The same benchmark category is used below to interpret the results of other regressions. For this category of respondent, increasing our indicator of religiosity by one standard deviation (+0.315) raises the desire to emigrate by 12.2 percentage points.[23] Note that the mean proportion of aspiring migrants equals 24.3% in the MENA region. Although positive, the effect of gender-egalitarian views is not significantly different from zero. Control variables are usually significant and have intuitive signs. In line with the literature, aspirations are higher for young, single men with higher education, with lower level of income per household member, and who have friends or relatives abroad. Interestingly, our quadratic specification identifies a hump-shaped effect of household income, in line with the mobility transition literature (see Zelinsky (1971)).

The rest of the table distinguishes between migration aspirations to OECD and to non-OECD destinations. Columns (5) and (6) reveal that cultural traits have insignificant impact on migration aspirations to non-OECD countries. On the contrary, low levels of religiosity is positively correlated at the 1% significance level with intentions to emigrate to OECD destinations, as shown in Column (4), and the effect is bigger compared to the results in Column (2). Increasing our indicator by one standard deviation raises the desire to emigrate to OECD destinations by 14.9 percentage points (for the benchmark category of respondent).The correlation between gender-egalitarian views is also bigger, however it still not precisely estimated. Aspirations to migrate to OECD destinations are even more influenced by education attainment and by the presence of network members abroad. Hence, Table 5.3 evidences that aspiring migrants from MENA countries self-select along cultural traits but only when they intend to migrate to an OECD destination; they exhibit lower levels of religiosity than those who do not intend to migrate.[24]

[23] Increasing the religiosity index implies reducing the level of religiosity

[24] In Table A12 in the Appendix we also test whether those results are mostly driven by a deviation from

Table 5.3: Benchmark regressions

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	Int Mig	Int Mig	Int Mig	Int Mig	Int Mig	Int Mig
Culture	Gend	Rel	Gend	Rel	Gend	Rel
Destination	All	All	OECD	OECD	non OECD	non OECD
Culture	0.051	0.389***	0.142	0.472***	-0.092	0.227
	(0.101)	(0.102)	(0.103)	(0.101)	(0.150)	(0.168)
Female	-0.658***	-0.598***	-0.714***	-0.701***	-0.588***	-0.411***
	(0.107)	(0.097)	(0.100)	(0.090)	(0.154)	(0.149)
Network	0.677***	0.615***	0.708***	0.676***	0.689***	0.537***
	(0.067)	(0.060)	(0.083)	(0.069)	(0.072)	(0.076)
Age	-0.296***	-0.261***	-0.326***	-0.305***	-0.256***	-0.188***
	(0.039)	(0.040)	(0.044)	(0.040)	(0.040)	(0.044)
Education	0.179***	0.113	0.253**	0.221*	0.084	-0.019
	(0.058)	(0.074)	(0.103)	(0.116)	(0.074)	(0.078)
Marital	-0.257***	-0.401***	-0.354***	-0.481***	-0.167***	-0.291***
	(0.055)	(0.088)	(0.058)	(0.076)	(0.061)	(0.108)
Children	-0.049	-0.021	-0.056	-0.039	-0.010	0.012
	(0.070)	(0.040)	(0.100)	(0.038)	(0.072)	(0.073)
Income	-0.088**	-0.075**	-0.046	-0.068*	-0.136**	-0.103***
	(0.036)	(0.032)	(0.042)	(0.041)	(0.053)	(0.030)
$Income^2$	0.005**	0.002***	0.003	0.002**	0.008**	0.003***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.003)	(0.001)
Observations	32805	34017	28937	30038	28065	28240
Partial R^2	0.11	0.11	0.17	0.18	0.11	0.10
Region f.e.	Yes	Yes	Yes	Yes	Yes	Yes
Year-Wave f.e.	Yes	Yes	Yes	Yes	Yes	Yes

Source: Author's calculations on Gallup Data. Std. errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Std. errors are clustered at the country level. The dependent variable is an intention to migrate dummy. OECD destinations: US, UK, France, Germany, Netherlands, Spain, Italy, Poland, Hungary, Sweden, Greece, Denmark, Israel, Canada, Australia, New Zealand, South Korea, Austria, Estonia, Finland, Japan, Mexico, Belgium, Turkey, Iceland, Ireland, Latvia, Norway, Portugal, Slovenia, Switzerland, Czech Rep.

Table 5.4: Results for intentions and plans to migrate

	(1)	(2)	(3)	(4)
Dep. Var.	Int Mig	Plan Mig	Int Mig	Plan Mig
Culture	Gend	Gend	Rel	Rel
Destination	OECD	OECD	OECD	OECD
Culture	0.142	-0.077	0.472***	0.771***
	(0.103)	(0.173)	(0.101)	(0.243)
Observations	28937	10834	30038	12199
Partial R^2	0.17	0.16	0.18	0.23
Region f.e.	Yes	Yes	Yes	Yes
Year-Wave f.e.	Yes	Yes	Yes	Yes

Source: Author's calculations on Gallup Data. Std. errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Std. errors are clustered at the country level. OECD destinations. Dep. Var.: intention to migrate (Int Mig); plan to migrate in the next 12 months (Plan Mig)

In Table 5.4, we check whether similar selection patterns apply to individual who have concrete migration plans, i.e. those who are taking concrete steps to leave their country within the next 12 months. Columns (1) and (3) report the results from Table 5.3 for migration aspirations. Relying on the same specification, Columns (2) and (4) provides the results for migration plans. Column (2) shows that the effect of gender-egalitarian views is insignificant. On the contrary, the effect of religiosity is highly significant and greater than for migration aspirations. We obtain a coefficient of 0.771, which means that increasing our indicator by one standard deviation (+0.315) raises the probability to have concrete migration plans by 24.3 percentage points. Note that the mean proportion of individuals taking steps to move within 12 months equals 1.9% and its standard deviation equals 13.9 percentage points. We thus find evidence of an effect of religiosity on migration plans, implying that emigration to OECD countries affects the distribution of cultural traits in the population left behind.

the average level of culture at the regional level. To do so, we conduct our analysis without fixed effects (or with country fixed effects only). Intending migrants always self-select on religiosity, and results are magnified when individuals deviate from the region/country mean.

5.4.2 Robustness Analysis

Robustness by subsample

In this section, we investigate whether the identified self-selection patterns vary by intended country of destination or by group of individuals.

Table 5.5: Robustness analysis by preferred destination countries

	(1)	(2)	(3)	(4)	(5)	(6)
Culture	Gend	Rel	Gend	Rel	Gend	Rel
Destination	EU	EU	US/Can	US/Can	Turk	Turk
Culture	0.014	0.408***	0.170	0.679***	-0.021	0.095
	(0.115)	(0.108)	(0.160)	(0.135)	(0.283)	(0.306)
Observations	25617	27210	24645	24840	11973	12031
Partial R^2	0.20	0.21	0.16	0.13	0.27	0.13
Region f.e.	Yes	Yes	Yes	Yes	Yes	Yes
Year-Wave f.e.	Yes	Yes	Yes	Yes	Yes	Yes

Source: Author's calculations on Gallup Data. Std. errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Std. errors are clustered at the country level. The dependent variable is an intention to migrate dummy. Col. (1) and (2): European destinations that are member states of the OECD; Col. (2) and (3): US and Canada; Col. (5) and (6) Turkey.

We begin by splitting the set of OECD destinations into three subsets of countries that are frequently reported as preferred destinations in the data, namely the European Union, North America (i.e. Canada and the United States) and Turkey. Separate fixed-effect logit regressions are used to explain migration aspirations to these three sets of countries.[25] Results are provided in Table 5.5. Columns (1), (3) and (5) confirm that the effect of gender-egalitarian views remains insignificant for all sets of destination. Columns (2) and (4) show that the effect of religiosity is highly significant when considering OECD, high-income destinations. We also notice that the intensity of the self-selection process is greater for individuals intending to migrate to North America (0.679) than for those who intend to migrate to Europe (0.408). On the contrary, we find no evidence of cultural selection towards Turkey.

[25] We disregard interdependencies between intended location decisions, which could be accounted for using a multinomial logit model.

Table 5.6: Gender views and migration aspirations towards OECD countries
Subsample Analysis

Subsamples	*Benchmark*	*Gender & Education*			*Age*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Men	Men HS	Men LS	15-30	31-45	46-65
Gender Att.	0.142	0.205	0.226	0.167	0.258**	-0.213	0.351
	(0.103)	(0.149)	(0.201)	(0.190)	(0.113)	(0.188)	(0.268)
Observations	28937	14425	5312	8678	13586	8657	4832
Partial R^2	0.17	0.16	0.18	0.15	0.18	0.16	0.14

	Religion			*Gender & Marital Status (S,M)*			
	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Muslim	Christian	Other	Men (S)	Men (M)	Women (S)	Women (M)
Gender Att.	0.198	-0.519	0.104	0.269	0.098	0.447***	-0.148
	(0.142)	(0.350)	(0.104)	(0.194)	(0.156)	(0.164)	(0.218)
Observations	21638	2008	5204	5691	8179	4900	7487
Partial R^2	0.19	0.11	0.16	0.15	0.13	0.15	0.17

	Place of Residence			*Employment Status*		
	(15)	(16)	(17)	(18)	(19)	(20)
	Rural	Town	City	Unemp	Employ	Out LF
Gender Att.	0.492**	0.212	-0.055	0.668	0.186	0.139
	(0.211)	(0.166)	(0.127)	(0.408)	(0.156)	(0.174)
Observations	6946	7304	12926	1347	12181	9589
Partial R^2	0.23	0.19	0.15	0.22	0.16	0.20

Source: Author's calculations on Gallup Data. Std. errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Std. errors are clustered at the country level. In each regression we control for a set of individual level characteristics, region and year fixed effects. The dependent variable is an intention to migrate dummy towards OECD destination countries. The benchmark results are available in column (1). We then perform a subsample analysis on the following characteristics: Gender and Education (Men (2), Men with more than 9 years of education (3), Men with less than 9 years of education (4)), Age groups (15-30 (5), 31-45 (6) and 46-65(7)), Religion group (Muslims (8), Christians (9) and Others (10)), Gender and Marital status group (Single Men (11), Married Men (12), Single Women (13) and Married Women (14)), Place of Residence (Rural area (15), Town or Village (16) and City or suburbs (17)) and Employment status (unemployed (18), employed (19) and out of the labor force (20))

Table 5.7: Religiosity and migration aspirations towards OECD countries
Subsample Analysis

Subsamples	*Benchmark*	*Gender & Education*			*Age*		
	(1)	(2) Men	(3) Men HS	(4) Men LS	(5) 15-30	(6) 31-45	(7) 46-65
Religiosity	0.472***	0.665***	0.688***	0.696***	0.420***	0.522***	0.512***
	(0.101)	(0.105)	(0.221)	(0.123)	(0.134)	(0.111)	(0.177)
Observations	30038	15031	5192	9515	13771	9260	5374
Partial R^2	0.18	0.17	0.22	0.15	0.16	0.16	0.14

	Religion			*Gender & Marital Status (S,M)*			
	(8) Muslim	(9) Christian	(10) Other	(11) Men (S)	(12) Men (M)	(13) Women (S)	(14) Women (M)
Religiosity	0.518***	0.180	0.511	0.604***	0.757***	0.122	0.380*
	(0.110)	(0.164)	(0.327)	(0.123)	(0.147)	(0.215)	(0.209)
Observations	22421	1952	5487	6171	8515	5441	7652
Partial R^2	0.18	0.10	0.13	0.12	0.14	0.16	0.15

	Place of Residence			*Employment Status*		
	(15) Rural	(16) Town	(17) City	(18) Unemp	(19) Employ	(20) Out LF
Religiosity	0.555**	0.272*	0.397***	0.306	0.433***	0.305
	(0.227)	(0.162)	(0.111)	(0.411)	(0.126)	(0.199)
Observations	6729	5619	12733	874	8389	6862
Partial R^2	0.24	0.19	0.15	0.22	0.17	0.24

Source: Author's calculations on Gallup Data. Std. errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Std. errors are clustered at the country level. In each regression we control for a set of individual level characteristics, region and year fixed effects. The dependent variable is an intention to migrate dummy towards OECD destination countries. The benchmark results are available in column (1). We then perform a subsample analysis on the following characteristics: Gender and Education (Men (2), Men with more than 9 years of education (3), Men with less than 9 years of education (4)), Age groups (15-30 (5), 31-45 (6) and 46-65(7)), Religion group (Muslims (8), Christians (9) and Others (10)), Gender and Marital status group (Single Men (11), Married Men (12), Single Women (13) and Married Women (14)), Place of Residence (Rural area (15), Town or Village (16) and City or suburbs (17)) and Employment status (unemployed (18), employed (19) and out of the labor force (20))

We now focus on migration aspirations to OECD destination countries and test whether the intensity of cultural selection varies by group of individuals. Kaestner and Malamud (2014) note that individual characteristics determines migrants and non-migrants in the Mexican context. We run logit regressions using different subsamples of individuals. We distinguish between *gender and education groups* (men, men high educated and men low educated), *age groups* (between 15-30, 31-45 and 46-65), *religious groups* (Muslims, Christians and other confessions), *gender and marital status* (single men, single women, married med, married women), *place of residence* (rural area, town or small village, city and suburbs) and *employment status* (unemployed, employed, out of the labor force). Table 5.6 shows the results for gender-egalitarian views. We identify a positive correlation between gender-egalitarian views and intentions to emigrate for individuals aged 15-30 (i.e., individuals exhibiting the greatest probability to realize their aspirations), for single women and for people living in rural areas. The coefficient is usually significant at the 5% threshold. On the contrary, the effect is insignificant for the other groups.

Results for religiosity are provided in Table 5.7. They confirm that less religious people are more prone to migrate towards OECD destination countries. Men self-selected more than the total population (0.655 against 0.472). The intensity of self-selection is homogenous across educaation groups. Self-selection also increases with age. Columns (8) to (10) show that cultural selection is totally explained by the behavior of the Muslim population. Columns (16) to (19) reveal that self-selection is stronger for employed individuals living in rural areas.

Robustness by using matched samples

We now investigate whether our results are driven by differences in the composition of the samples of aspiring migrants and non-migrants, and whether the threat of biased estimation due to unobservables can be mitigated. In line with comparisons between treated and control groups, we use the *Mahalanobis Metric Matching* technique to identify samples of aspiring migrants and non-migrants that are balanced in terms of covariates. We run logit regressions using a sample of individuals that are similar in terms of observed covariates (apart from their cultural traits), and potentially similar in terms of unobservables if the latter are correlated with observed characteristics. The matching procedure minimizes the Mahalanobis metric.

For each covariate x, we compute the normalized difference:[26]

$$\Delta_x = (\bar{x}_{Mig} - \bar{x}_{SNon}) \left(\frac{s^2_{x,Mig} + s^2_{x,Non}}{2} \right)^{-1/2}$$

where the difference between the mean value of the covariate for aspiring migrants and non-migrants, $\bar{x}_{Mig} - \bar{x}_{Non}$, is divided by the mean of the standard deviations of the covariate over the whole sample ($s_{x,Mig}$ and $s_{x,Non}$).

Results of the matching technique are described in Table A13 in the Appendix. For each sample, we report the difference in terms of covariates before and after the matching procedure. Before matching, the distribution of covariates is unbalanced for both samples; differences in characteristics are always statistically different from zero. On the contrary, the matching technique allows generating a matched sample exhibiting a balanced distribution of covariates. After matching, the only variable along which aspiring migrants and non-migrants exhibit statistically different outcomes is the level of income per household member.

Table 5.8: Analysis on the Matched Sample

	(1)	(2)	(3)	(4)	(5)	(6)
Culture	Gend	Rel	Gend	Rel	Gend	Rel
Destination	All	All	OECD	OECD	non OECD	non OECD
Culture	0.078	0.333***	0.125	0.423***	-0.014	0.188
	(0.129)	(0.107)	(0.135)	(0.115)	(0.160)	(0.185)
Observations	11184	12100	8594	9442	8145	8320
Partial R^2	0.04	0.04	0.09	0.08	0.10	0.10
Region f.e.	Yes	Yes	Yes	Yes	Yes	Yes
Year-Wave f.e.	Yes	Yes	Yes	Yes	Yes	Yes

Source: Author's calculations on Gallup Data. Std. errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Std. errors are clustered at the country level. The dependent variable is an intention to migrate dummy. OECD destinations: same as Table 3.

Table 5.8 provides the results of the fixed-effect logit regressions using the matched samples; they can be easily compared with those of Table 5.3 for the non-matched samples. All conclusions of the benchmark regressions hold when using the matched samples. Columns

[26]Since some individuals did not answer all questions on religiosity and gender-egalitarian views, we decide to perform a specific matching procedure for each regression.

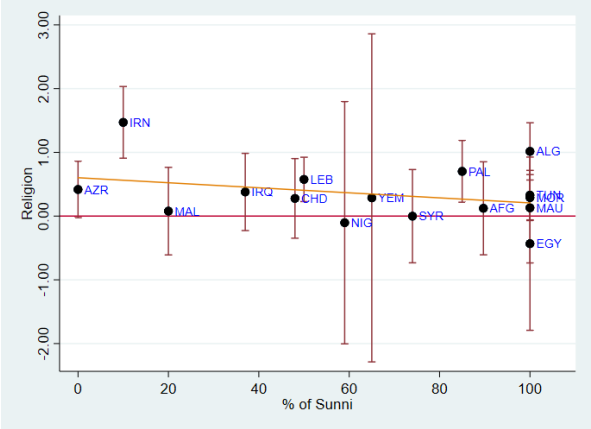
(1) and (2) confirm that aspiring migrants self-select in terms of religiosity, but not along gender-egalitarian views. Column (4) and (6) confirm that the results are driven by migration aspirations to OECD destination countries only. The coefficient of religiosity equals 0.423 and is significant at the 1% threshold; it is almost identical to that of Table 5.3. Increasing the indicator of religiosity by one standard deviation raises the desire to emigrate 13.4 percentage points.[27] The slight decrease in the estimated coefficient suggests that the matching technique allows reducing an upward bias due to omitted variables.

Heterogeneity across countries

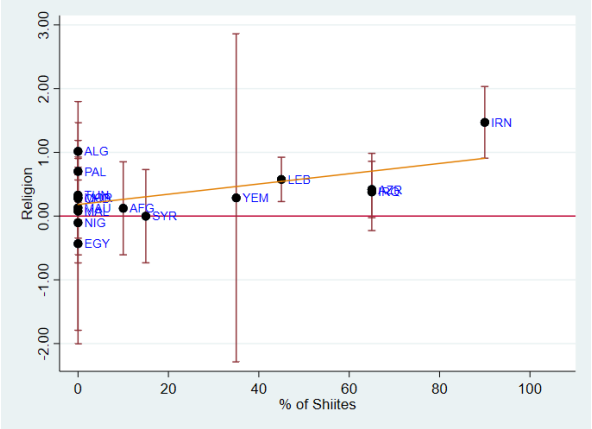
We now explore whether cultural selection varies across countries and with country-specific characteristics. The set of country characteristics includes the shares of Sunnis and Shiites among the Muslim population, the log of GDP per capita, two indicators of institutional quality, and the share of native citizens from the same origin country living in a OECD country member state, as a proxy of network size abroad. Figure 5.3 plots the coefficient of religiosity obtained from country-specific regressions against the level of country characteristics; it also shows the linear trend between them. It evidences large variations across countries, both in the size of the estimated coefficient and in its confidence interval. Such variations can be explained by differences in sample size or by more structural factors. However, the correlation between the estimated coefficients and country characteristics is always insignificant, except when considering the share of Shiites in the Muslim population. In the latter case, the correlation is positive (0.539). The same exercise is conducted for gender-egalitarian attitudes in Figure 5.4. It reveals a negative and significant correlation between the estimated effect of culture and two country characteristics, namely the share of Sunnis in the Muslim population (-0.521) and the control of corruption (-0.583).

[27] As before, this effect on migration aspirations is computed for the benchmark group, which consists of male individuals without friends or relatives abroad, with college education, married with children, between 24 to 35 years old, and with an average level of religiosity (0.229).

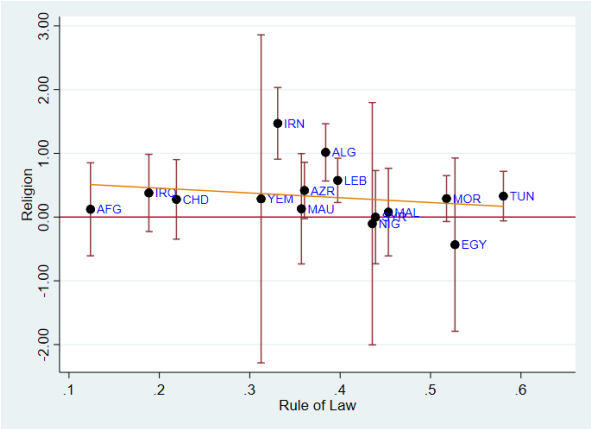
Figure 5.3: Religiosity: estimated $\hat{\beta}$ and its 95% interval of confidence by country



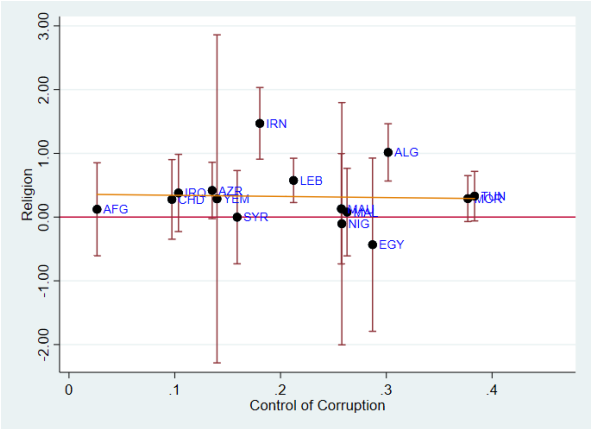
(a) Sunnis - (Corr = −0.304)



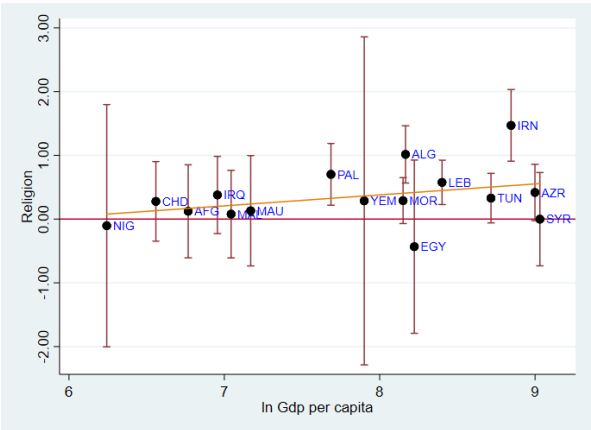
(b) Shiites - (Corr = 0.539**)



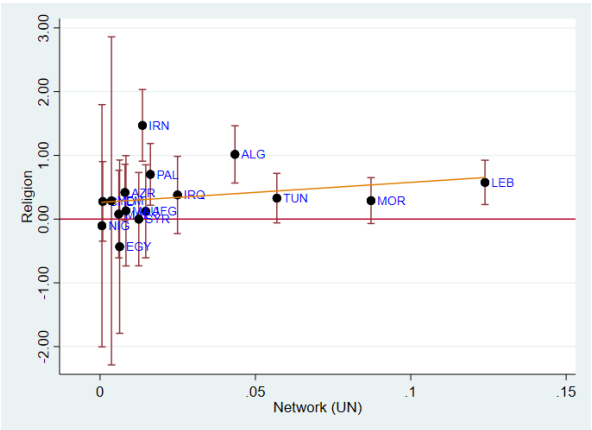
(c) Rule of Law - (Corr = −0.211)



(d) Control of Corruption - (Corr = −0.041)

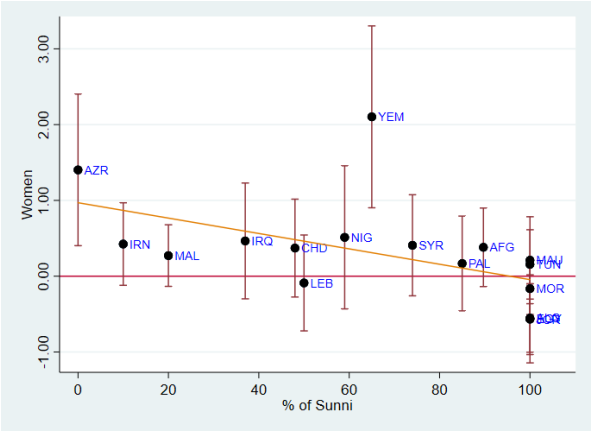


(e) ln GDP - (Corr = 0.349)

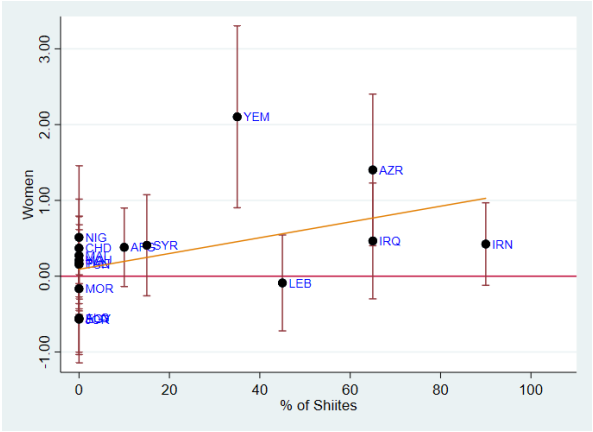


(f) Network - (Corr = 0.245)

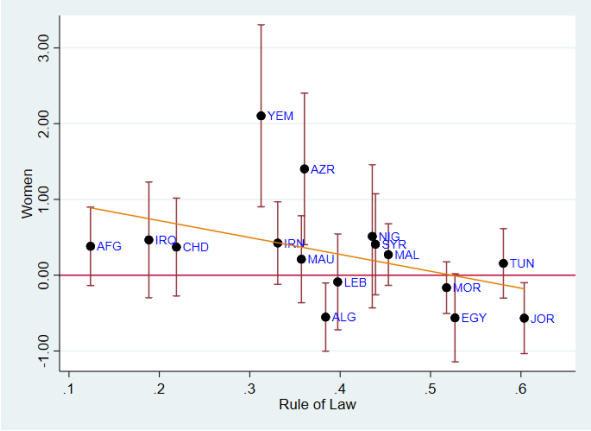
Figure 5.4: Gender-egalitarian views: estimated $\hat{\beta}$ and its 95% interval of confidence by country)



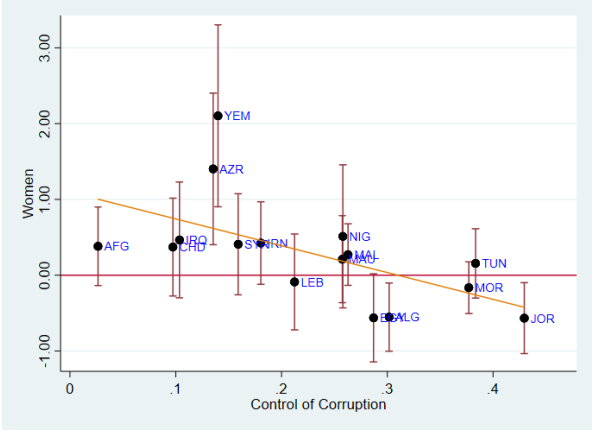
(a) Sunnis - (Corr = -0.521^{**})



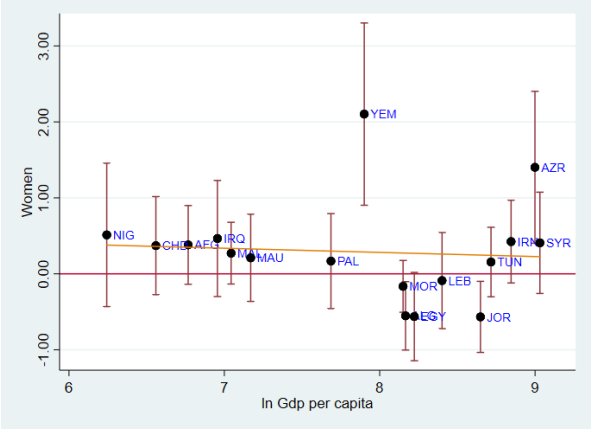
(b) Shiites - (Corr = 0.457^{*})



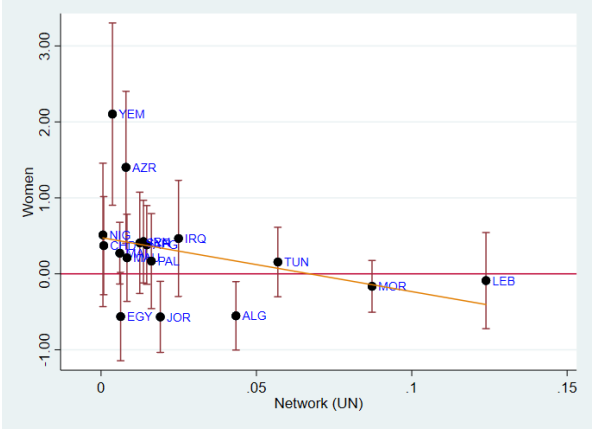
(c) Rule of Law - (Corr= -0.437^{*})



(d) Control of Corruption - (Corr = -0.583^{**})



(e) ln GDP - (Corr= -0.074)



(f) Network - (Corr= -0.359)

Table 5.9: Adding interactions with macroeconomic variables

	(1)	(2)	(3)	(4)	(5)	(6)
Destination	OECD	OECD	OECD	OECD	OECD	OECD
Interaction	Sunnis	Shiites	Cont. Corr.	Rule Law	ln GDP	Net.
Religiosity	0.589**	0.356**	0.490**	0.773**	0.373**	0.461***
	(0.242)	(0.138)	(0.219)	(0.333)	(0.178)	(0.151)
Interaction	-0.002	0.006	-0.150	-0.797	0.000	0.006
	(0.003)	(0.004)	(0.819)	(0.687)	(0.000)	(0.017)
Observations	30038	30038	27141	27141	30038	30038
Partial R^2	0.18	0.18	0.18	0.18	0.18	0.18
Gender Att.	0.764***	0.024	0.633**	0.639	0.140	0.233*
	(0.183)	(0.104)	(0.313)	(0.392)	(0.155)	(0.134)
Interaction	-0.009***	0.009***	-1.918*	-1.199	0.000	-0.029
	(0.002)	(0.003)	(1.029)	(0.827)	(0.000)	(0.020)
Observations	28937	28937	27469	27469	28937	28937
Partial R^2	0.17	0.17	0.18	0.18	0.17	0.17

Source: Author's calculations on Gallup Data. Std. errors in parentheses;*** p<0.01, ** p<0.05, * p<0.1. The dependent variable is an intention to migrate dummy. We control for gender, network, age cohort, level of education,marital status, presence of children, income, income squared, year-wave and region f.e. Std. errors are clustered at the country level. OECD destinations: same as Table 3

To generalize this descriptive analysis of correlations, we run regressions accounting for interactions between country characteristics and cultural traits. Results are provided in Table 5.9. Overall, the interaction between progressive views on religiosity and country characteristics is never significant, even when considering the interaction with the share of Shiites in the Muslim population. In addition, the effect of religiosity remains significant in all specifications. This reinforces our conclusion that aspiring migrants from virtually all MENA countries self-select along religiosity levels. As far as gender-egalitarian attitudes are concerned, the results in Column (1) suggest that aspiring migrants from countries with a Sunni majority have less gender-egalitarian views. This is confirmed in Column (2) which reports a positive coefficient for the interaction between cultural traits and the share of Shiites. Column (3) shows that selection along gender-egalitarian views also becomes significant when controlling for a measure of control of corruption in the origin country. This suggests that improving institutions reduces self-selection on gender-egalitarian attitudes.

Effect of the Arab Spring

We finally explore whether the link between cultural traits and migration has been affected by the Arab Spring. Since most questions on religiosity and gender-egalitarian views are asked from 2007 to 2011, we use the only proxy for religiosity that is available in all waves of the GWP (2007-2016). The question is: Is religion an important part of your daily life? Responses to this question are highly correlated with the synthetic indicator of religiosity resulting from our PCA, as it is shown in Table A6 in the Appendix.

We run fixed-effect logit regressions accounting for the interactions between a post-Arab Spring dummy (equal to one for the years 2011 to 2016) and religiosity. Results for the full sample are provided in Column (1) of Table 5.10.

The Arab Spring affected MENA countries differently. In countries like Tunisia, Egypt and Yemen, the government was overthrown after turmoils and riots. Algeria experienced two years of major protests, while the Arab Spring phenomenon lead Syria into a civil war, with more than 400,000 casualities and high numbers of refugees in the neighboring countries. The uprisings in these countries was triggered by the high level of unemployment (particularly for the young), by the persistence of economic inequality and by political authoritarianism.[28]

[28]Economic and political frustrations of the aspiring youth played an important role (Boughzala and Romdhane (2017)). In Syria authoritarian political system and the mismanagement of the economy led to uprisings, that developed into a civil war eventually drawing both regional and international external armed interventions (Safadi and Neaime (2017)). The possible persistence of authoritarianism in those countries could be related to cultural factors (Elbadawi and Makdisi (2017)).

Table 5.10: Cultural selection before and after the Arab Spring

	(1)	(2)	(3)	(4)	(5)	(6)
Destination	OECD	OECD	OECD	OECD	OECD	OECD
Sample	Full	Full	Full	Matched	Matched	Matched
Countries	All	Main Ins.	Others	All	Main Ins.	Others
Religiosity	0.473***	0.560***	0.409***	0.446***	0.559***	0.396***
	(0.080)	(0.088)	(0.116)	(0.062)	(0.108)	(0.082)
Interaction	-0.110	-0.475**	0.081	-0.149	-0.425*	0.010
	(0.177)	(0.215)	(0.203)	(0.173)	(0.244)	(0.228)
Observations	100,544	34,914	65,630	41,801	14,328	27,473
Region f.e.	Yes	Yes	Yes	Yes	Yes	Yes
Year-Wave f.e.	Yes	Yes	Yes	Yes	Yes	Yes
Partial R^2	0.139	0.170	0.127	0.142	0.173	0.131

Source: Author's calculations on Gallup Data. Std. errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is an intention to migrate dummy. We control for gender, network, age cohort, level of education, marital status, presence of income squared, the interaction terms with individual controls and Arab Spring, children, income, and region f.e.. Std. errors are clustered at the country level. Sample legend: Full Sample (FS); Matched on After Arab Spring (Matched). OECD destinations: same as Table 3

The turbulent events increased aversion to the use of violence by civilians and military armies.[29] We refer to these five countries (Algeria, Egypt, Syria, Tunisia and Yemen) as the group of *Main Insurgents*. We then run a subsample analysis of the effect of Arab Spring on the *Main Insurgents* and on the other countries. Intuitively, we should expect a greater effect for the *Main Insurgents*. Columns (1) to (3) use the full sample; Columns (4) to (6) use matched samples, with matching based on the post-Arab Spring period.[30] Results of the matching are available in Table A14 and minimize possible biases due to the unbalanced distribution of covariates and unobservables.

In all specifications, selection by religiosity is always positive and significant. Although the Arab Spring has not affected the intensity of cultural selection in the less affected countries, it has reduced it in the *Main Insurgent* countries. The latter findings are highly robust to the use of the matching technique. In other words, the Arab Spring has increased the relative religiosity of aspiring migrants. The data do not allow us to investigate whether the degree of cultural selection has changed after the recent conflicts and political unrest in the Middle-East. Given similarities with the Arab Spring period, it is however plausible that

[29]GWP includes a few questions on attitudes about violence towards civilians. Compared to other countries, respondents from Tunisia, Egypt, Yemen, Algeria and Syria are less prone to justify the killing of civilians by military or individuals after the Arab Spring.

[30]We match individual before and after the Arab Spring.

the 2015 wave of asylum seekers is less culturally selected than previous migration waves.

5.5 Discussion and policy issues

Our empirical results indicate that individuals who intend to emigrate to OECD, high-income countries exhibit significantly lower levels of religiosity than the rest of the population. They also share more gender-egalitarian views, although this effect only holds among the young. Potentially, these results have implications from the point of view of both origin and destination countries.

From the *point of view of destination countries*, selection on religiosity and gender-egalitarian attitudes implies that the cultural distance between migrants and host country citizens is smaller than between the country populations. This means easier integration to the host country labor market and social norms. Selective migration from MENA to high income OECD countries should be less of a concern from the point of view of the OECD member states. On the one hand, *informing public opinion in this regard* might influence attitudes towards immigration and discrimination practices; this seems a relevant policy recommendation. On the other hand, a key finding of our analysis is that the aggregate effect of cultural selection should not be overestimated.

The left panel of Figure 5.5(a) compares the average level of religiosity of OECD natives (vertical line at 0.623) with the average level of intending stayers in the MENA, of intending migrants towards OECD destinations, and of current migrants from MENA countries in the OECD (extracted from the GWP database). When considering all respondents, the religiosity index of intending migrants (0.291) is 13% higher than that of intending stayers (0.257). Remember, zero corresponds to the maximal religiosity level. Hence, self-selection along religiosity levels reduces the gap between MENA and OECD countries by 9% only. Similar findings are found when comparing young intended migrants and non-migrants. When considering highly educated respondents, the average religiosity index is slightly greater (0.301) but cultural differences between intended migrants and non-migrants become negligible.[31] The GWP data also enable us to compare actual migrants from MENA countries (those who have already migrated) with OECD native citizens. Figure 5.5(a) shows that the religiosity index of actual migrants is much closer to that of natives, and this is especially true for older migrants and for the highly educated. The gap between actual and intended migrants can

[31] We do not identify a strong and significantly different effect of religiosity and gender-egalitarian views on migration aspirations of highly educated individuals (see Tables 5.7 and 5.6). However, this effect is dominated by other distributional characteristics of the highly educated population.

be due to several reasons: (i) cultural selection in the realization of migration aspirations (in line with the results of Table 5.4), (ii) a gradual decline in cultural selection over time (in line with our findings about the Arab Spring, see Table 5.10), (iii) a sign of cultural assimilation abroad, (iv) estimation biases due to the under representation of the foreign-born population in the survey (the GWP database only includes 282 immigrants from the MENA in the OECD countries).

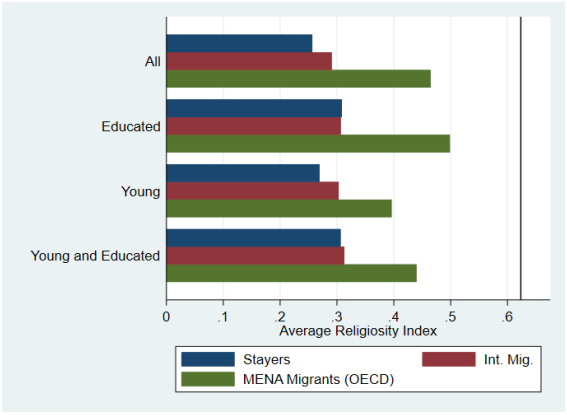
As for gender-egalitarian views, data for OECD countries are not available but we estimate that the reference level is close to unity. Figure 5.5(b) shows that intended migrants are slightly more progressive, an effect that is mostly driven by the young population (in line with Table 5.6). On average, selection on gender-egalitarian views is small. In sum, despite cultural selection, immigrants from the MENA countries exhibit markedly different cultural traits compared to OECD natives.

Another important finding is that the Arab Spring has decreased the strength of cultural selection. Although the external validity of our results cannot be assessed, this result suggests that conflicts and political turmoil in developing countries are likely to increase not only migration pressures to OECD countries but also the average cultural distance between immigrants and host-country citizens. Migration peaks and cultural distance are two sources of tension and negative attitudes towards immigrants in the OECD member states. Hence, we see cultural selection as another argument justifying coordinated interventions to bring peace, security and stability in countries at risk.

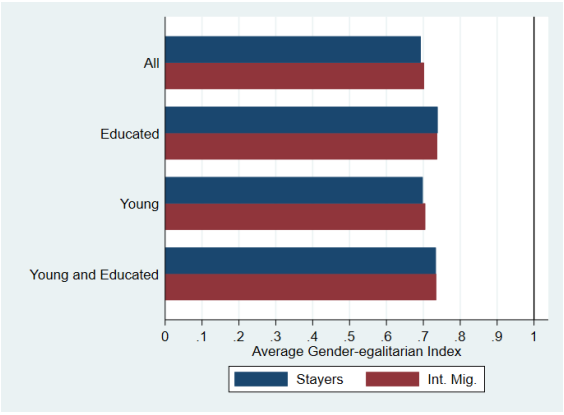
From *the point of view of the home country*, the distribution of cultural traits in the population left behind tends to be skewed towards more religiosity and less gender-egalitarian attitudes. Emigration to OECD countries impacts the distribution of cultural traits among those left behind, with potential implications for modernization, growth and democracy. On this basis, it could be argued that emigration should be combated if the home-country government targets to achieve a higher level of economic development. Our results do not support this view. Religiosity and gender-egalitarian attitudes are correlated with other observed and unobserved characteristics that affect migration aspirations. Figure 5.5(c) compares the observed average indices of religiosity and gender-egalitarian view of the current MENA population (in blue), with those obtained if all intended migration had left their country (in red), or if all young intended migrants had left (in green). Given the proportion of intended migrants towards OECD countries (12,3% on average) and the small cultural differences between groups, the average cultural traits of the population left behind hardly changes under these two counterfactuals. In sum, despite cultural selection, emigration from the MENA countries is unlikely to induce negative effects on modernization, growth and democracy. On

the contrary, emigration towards OECD countries could even reverse the effect on average cultural traits of the selection effect if migrants abroad transfer more progressive norms and beliefs to their home country, as argued in Rapoport *et al.* (2018) and Bouoiyour and Miftah (2017).

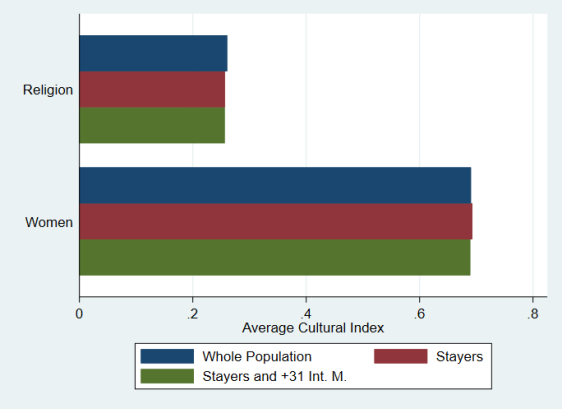
Figure 5: Average cultural indices by group of respondents



(a) Religiosity Index (migrants vs non-migrants)



(b) Gender-Egal. Index (migrants vs non-migrants)



(c) Cultural traits in the MENA

The graph shows the mean level of religiosity and gender-egalitarian views of the intending stayers (Stayers), intending migrants towards OECD (Int. Mig) and the migrants already in the OECD (MENA Migrants (OECD)) on panels (a) and (b). The vertical black line is the mean level of religiosity of natives in the OECD on panel (a), and a society sharing fully gender-egalitarian views on panel (b). The means are computed for four groups: the overall population (All), the population between 15 and 30 (Young), the population with at least 9 year of education (Educated) and the young population with high education (Young and Educated). Panel (c) shows the mean level of religiosity and gender-egalitarian views of the whole population, of intending stayers, and of the whole population minus young intending migrants. Author's calculations based on Gallup World Poll and World Bank Data.

5.6 Conclusions

In this paper, we use a unique database on migration aspirations, opinions and beliefs in the MENA countries to test whether migrants positively self-select by culture traits. We conduct fixed-effect logit regressions using the full sample or using matched samples of aspiring migrants and non-migrants. We find that aspiring migrants in general, and those who have concrete migration plans to migrate in particular, are culturally selected, and that this selection on cultural traits depends on the type of preferred destination. Intended migrants to OECD, high-income countries exhibit lower level of religiosity. This result is robust across gender, age groups and education levels. As far as attitude towards women's rights are concerned, aspiring migrants have more gender-egalitarian views in the age group 15-30, when they are single women, or when they originate from countries with a Sunni minority. Finally, we find a robust effect of the Arab Spring on the intensity of cultural selection only in countries highly impacted by the Arab Spring. In these countries, the Arab Spring has decreased the degree of cultural selection.

Given the high correlation between aspirations and actual migration flows, these results have implications from the point of view of both origin and destination countries. From the point of view of destination countries, cultural selection implies that the cultural distance between migrants and host country citizens is smaller than between the country populations. From the point of view of the home country, the distribution of cultural traits in the population left behind tends to be skewed towards more religiosity and less gender-egalitarian attitudes, with potential implications for modernization, growth and democracy. However, the aggregate effects of cultural selection are limited, albeit non negligible in post-conflict periods. Selective emigration has limited effects on the distribution of cultural traits in the MENA countries, and on the cultural distance between natives and immigrants in the OECD countries. To quantify the impact of emigration on culture, these effects should be compared with those linked to ex-post transfers of norms and beliefs from diasporas abroad to those left behind. Further research on such transfers of cultural norms and beliefs is on our agenda.

5.7 Appendix

5.7.1 List of preferred destinations by country of origin

Table A1: Top-5 preferred OECD destinations

Origin	1st best	2nd best	3rd best	4th best	5th best
Afghanistan	United States	Germany	Canada	UK	Turkey
Algeria	France	Canada	UK	Italy	Spain
Azerbaijan	Turkey	Germany	United States	UK	France
Chad	United States	France	Canada	UK	Japan
Egypt	United States	Italy	France	Germany	UK
Iran	Canada	United States	Germany	Australia	UK
Iraq	Sweden	United States	UK	Germany	Canada
Jordan	United States	Canada	Germany	Sweden	Turkey
Lebanon	Canada	United States	Australia	France	Germany
Mali	France	United States	Spain	Canada	Germany
Mauritania	France	United States	Spain	Canada	Germany
Morocco	France	Spain	Italy	Canada	United States
Niger	United States	France	Belgium	Canada	Germany
W.Bank and Gaza	United States	Sweden	Canada	Norway	Germany
Syria	United States	Canada	Sweden	Germany	Turkey
Tunisia	France	Italy	Germany	Canada	United States
Yemen	United States	UK	Turkey	Canada	Germany

Table A2: Top-5 preferred non-OECD destinations

Origin	1st best	2nd best	3rd best	4th best	5th best
Afghanistan	Iran	Saudi Arabia	Pakistan	UAE	Tajikistan
Algeria	UAE	Saudi Arabia	Egypt	Other	Qatar
Azerbaijan	Russia	Ukraine	Iran	Egypt	India
Chad	Nigeria	Cameoon	Saudi Arabia	Sudan	Senegal
Egypt	Saudi Arabia	UAE	Kuwait	Libya	Jordan
Iran	UAE	Other	Malaysia	Jordan	Russia
Iraq	UAE	Syria	Lebanon	Other	Egypt
Jordan	UAE	Saudi Arabia	Kuwait	Qatar	Palestine
Lebanon	UAE	Saudi Arabia	Qatar	Kuwait	South Africa
Mali	Ivory Coast	Angola	China	Equatorial Guinea	South Africa
Mauritania	UAE	Morocco	Saudi Arabia	Senegal	Qatar
Morocco	Saudi Arabia	UAE	Russia	Egypt	Qatar
Niger	Nigeria	Ivory Coast	Ghana	Libya	Togo
W.Bank and Gaza	UAE	Saudi Arabia	Jordan	Egypt	Other
Syria	UAE	Saudi Arabia	Qatar	Kuwait	Lebanon
Tunisia	UAE	Saudi Arabia	Libya	Other	Qatar
Yemen	Saudi Arabia	UAE	Qatar	Egypt	Other

5.7.2 Aggregate cultural indices

Principal Component Analysis (PCA)

Figure A1: Principal Component Analysis - Loadings plot

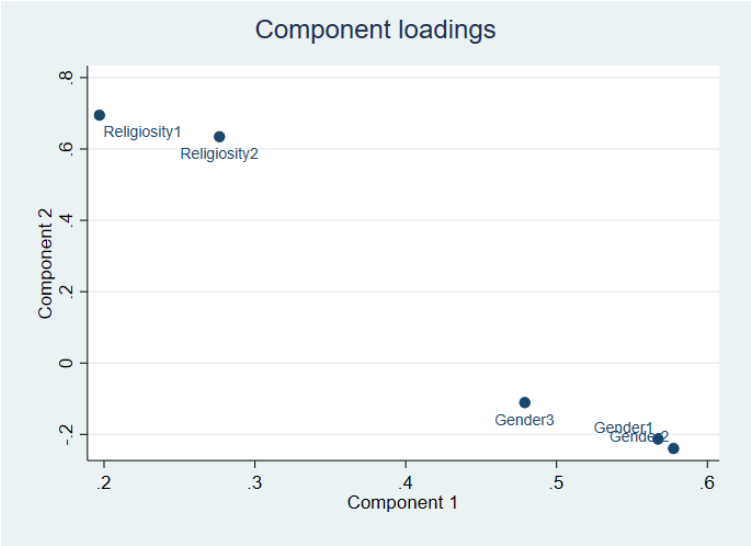


Table A3: Principal component eigenvectors

Variable	Comp1	Comp2	Comp3	Comp4	Comp5
Religiosity 1	0.1969	0.6946	0.1626	0.6719	-0.0286
Religiosity 2	0.2765	0.6344	-0.2552	-0.6752	-0.0015
Gender-egal. 1	0.5673	-0.2129	-0.3819	0.1751	0.6756
Gender-egal. 2	0.5776	-0.2396	-0.2516	0.1082	-0.7308
Gender-egal. 3	0.4789	-0.1108	0.8362	-0.2242	0.0938

Factor Analysis

Figure A2: Factor Analysis - Loadings plot

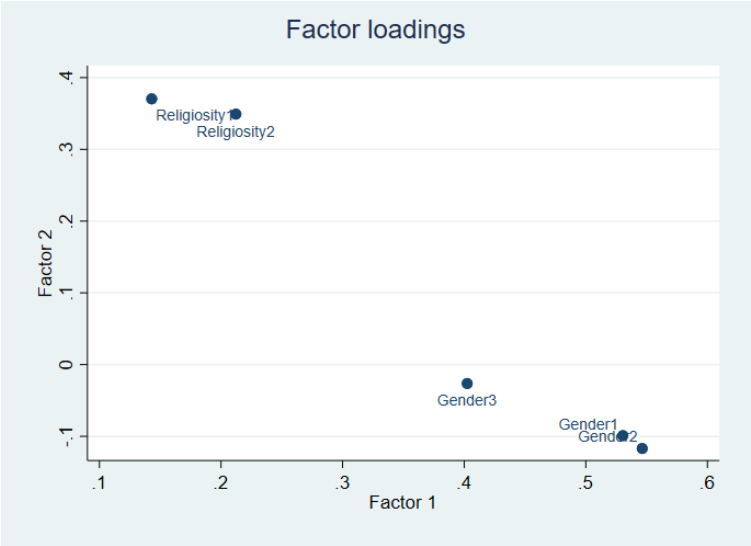


Table A4: Factor loadings and unique variance

Variable	Factor 1	Factor 2	Uniqueness
Religiosity 1	0.1430	0.3704	0.8424
Religiosity 2	0.2122	0.3491	0.8331
Gender-egal. 1	0.5305	-0.0990	0.7088
Gender-egal. 2	0.5464	-0.1170	0.6878
Gender-egal. 3	0.4025	-0.0263	0.8373

5.7.3 Descriptive Statistics

Table A5: Descriptive statistics & Robustness Analysis

Variable	Obs	Mean	s.d.	Min	Max
Demographic characteristics					
Migration desire	60284	0.243	0.429	0	1
Plan to migrate	60284	0.019	0.139	0	1
Female	60284	0.488	0.499	0	1
Network	60284	0.231	0.422	0	1
Age	60284	33.978	12.729	16	65
Education	60284	0.594	0.491	0	1
Marital	60284	0.586	0.492	0	1
Children	60284	0.718	0.450	0	1
Income HH	58600	1.23	1.613	0	111.539
Cultural characteristics					
Relig. (PCA)	34906	0.259	0.315	0	1
Religiosity (FA)	34906	0.262	0.317	0	1
Religiosity (MCA)	34906	0.212	0.261	0	1
Religiosity (KPCA,std)	34090	0.393	0.439	0	1
Religiosity (AVG)	34906	0.267	0.320	0	1
Gender-egal. (PCA)	33908	0.695	0.326	0	1
Gender-egal. (FA)	33908	0.702	0.325	0	1
Gender-egal. (MCA)	33908	0 .657	0.229	0	1
Gender-egal. (KPCA, std)	33148	0.495	0 .241	0	1
Gender-egal. (AVG)	33908	0.686	0.328	0	1

Authors' elaboration on Gallup data.

5.7.4 Robustness Analysis

Table A6: Pearson correlation between aggregate indices and GWP responses

Indicator	Religiosity (PCA)	Gender-egal. (PCA)
Religiosity 1	0.6880*	0.0671*
Religiosity 2	0.8784*	0.0790*
Gender-egal. 1	0.0909*	0.7446*
Gender-egal. 2	0.0526*	0.7592*
Gender-egal. 3	0.0702*	0.6992*

Source: Author's calculations on Gallup Data. * p<0.01.

Table A7: Pearson correlation across variables

	Relig. (PCA)	Gend. (PCA)	Fem	Netw	Age	Educ	Marital	Child	Inc
Relig. (PCA)	1.00								
Gend. (PCA)	0.0932*	1.00							
Female	0.2033*	0.2690*	1.00						
Network	-0.0104	0.0569*	-0.0075	1.00					
Age	-0.0468*	-0.0263*	-0.0395*	-0.0481*	1.00				
Education	0.1226*	0.1354*	-0.0840*	0.0928*	-0.1920*	1.00			
Marital	-0.0754*	-0.0545*	0.0170*	-0.0684*	0.4401*	-0.1736*	1.00		
Children	-0.1329*	-0.0969*	-0.0037	-0.0256*	-0.0656*	-0.1152*	0.2150*	1.00	
Income HH	0.0814*	0.0965*	-0.0173*	0.1019*	-0.0083	0.1965*	-0.0601*	-0.1270*	1.00

Notes: * p<0.01. Source: Authors' elaboration on Gallup data.

Table A8: Pearson correlation between aggregate indices (religiosity)

Indicator	PCA	Relig. 1	Relig. 2	FA	MCA	KPCA	AVG
Relig. 1	0.69*	1.00					
Relig. 2	0.8784*	0.2576*	1.00				
Relig. (FA)	0.9999*	0.6798*	0.8837*	1.00			
Relig. (MCA)	0.9275*	0.8095*	0.7019*	0.9248*	1.00		
Relig. (KPCA)	0.9217*	0.3565*	0.9947*	0.9260*	0.7672*	1.00	
Relig. (AVG)	0.9995*	0.6644*	0.8933*	0.9998*	0.9195*	0.9337*	1.00

Source: Author's calculations on Gallup Data. * p<0.01. The table shows correlations across the two questions on religion and the different way to aggregate them: PCA (Principal Component Analysis), FA (Factor Analysis), MCA (Multiple Correspondence Analysis), KPCA (Kernel Principal Component Analysis) and AVG (same weight to each question).

Table A9: Pearson correlation between aggregate indices (gender-egal. views)

Indicator	PCA	Gend. 1	Gend. 2	Gend. 3	FA	MCA	KPCA	AVG
Gend. 1	0.7446*	1.00						
Gend. 2	0.7592*	0.3834*	1.00					
Gend. 3	0.6992*	0.2523*	0.2912*	1.00				
Gend (FA)	0.9994*	0.7558*	0.7723*	0.6734*	1.00			
Gend. (MCA)	0.9729*	0.7266*	0.7654*	0.6512*	0.9738*	1.00		
Gend. (KPCA)	0.7416*	0.4349*	0.5193*	0.6830*	0.7333*	0.7072*	1.00	
Gend. (AVG)	0.9982*	0.7234*	0.7356*	0.7402*	0.9955*	0.9687*	0.7536*	1.00

Source: Author's calculations on Gallup Data. * p<0.01. The table shows correlations across the three questions on gender attitudes and the different way to aggregate them: PCA (Principal Component Analysis), FA (Factor Analysis), MCA (Multiple Correspondence Analysis), KPCA (Kernel Principal Component Analysis) and AVG (same weight to each question).

Table A10: Benchmark Regressions with Gulf Countries

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	Int Mig	Int Mig	Int Mig	Int Mig	Int Mig	Int Mig
Culture	Gend	Rel	Gend	Rel	Gend	Rel
Destination	All	All	OECD	OECD	non OECD	non OECD
Culture	0.034	0.402***	0.123	0.488***	-0.101	0.233
	(0.099)	(0.101)	(0.103)	(0.102)	(0.147)	(0.167)
Fem	-0.644***	-0.590***	-0.704***	-0.692***	-0.574***	-0.405***
	(0.104)	(0.096)	(0.098)	(0.090)	(0.150)	(0.148)
Net	0.665***	0.624***	0.698***	0.684***	0.676***	0.545***
	(0.066)	(0.060)	(0.084)	(0.069)	(0.069)	(0.077)
Age	-0.289***	-0.259***	-0.318***	-0.305***	-0.251***	-0.186***
	(0.039)	(0.040)	(0.044)	(0.039)	(0.040)	(0.045)
edu	0.164***	0.104	0.253**	0.208*	0.053	-0.019
	(0.057)	(0.074)	(0.102)	(0.118)	(0.073)	(0.076)
Marital	-0.255***	-0.401***	-0.349***	-0.477***	-0.167***	-0.294***
	(0.053)	(0.087)	(0.057)	(0.075)	(0.059)	(0.107)
Child	-0.040	-0.013	-0.049	-0.032	-0.004	0.022
	(0.070)	(0.039)	(0.099)	(0.037)	(0.071)	(0.072)
Income	-0.005	-0.052**	0.001	-0.038	-0.016	-0.093***
	(0.018)	(0.021)	(0.022)	(0.027)	(0.023)	(0.036)
Inc2	0.000	0.001*	-0.000	0.001	0.000	0.001***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)
Observations	35106	35257	31162	31234	30028	29446
Partial R^2	0.12	0.11	0.18	0.18	0.12	0.11
Region f.e.	Yes	Yes	Yes	Yes	Yes	Yes
Year-Wave f.e.	Yes	Yes	Yes	Yes	Yes	Yes

Source: Author's calculations on Gallup Data. Std. errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Std. errors are clustered at the country level. The dependent variable is an intention to migrate dummy. The pool of origin countries is composed by the 17 countries in our benchmark analysis, United Arab Emirates and Kuwait. OECD destinations: US, UK, France, Germany, Netherlands, Spain, Italy, Poland, Hungary, Sweden, Greece, Denmark, Israel, Canada, Australia, New Zealand, South Korea, Austria, Estonia, Finland, Japan, Mexico, Belgium, Turkey, Iceland, Ireland, Latvia, Norway, Portugal, Slovenia, Switzerland, Czech Rep.

Table A11: Logit regressions - Robustness to different indicators 2007-2011)
Dependent = Migration aspirations to OECD/non OECD destinations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Destination	OECD	OECD	OECD	OECD	OECD	Non OECD	Non OECD	Non OECD	Non OECD	Non OECD
Aggregation	PCA	FA	MCA	KPCA	AVG	PCA	FA	MCA	KPCA	AVG
Religiosity index	0.471***	0.465***	0.634***	0.216***	0.453***	0.229	0.224	0.366**	0.061	0.215
	(0.102)	(0.101)	(0.111)	(0.068)	(0.100)	(0.167)	(0.167)	(0.184)	(0.136)	(0.167)
Observations	30095	30095	30095	29385	30095	28297	28297	28297	27606	28297
Partial R^2	0.17	0.17	0.18	0.18	0.17	0.10	0.10	0.10	0.10	0.10
Gender att. index	0.150	0.153	0.318*	0.147	0.143	-0.091	-0.090	-0.037	0.027	-0.092
	(0.103)	(0.103)	(0.173)	(0.124)	(0.104)	(0.149)	(0.148)	(0.198)	(0.124)	(0.151)
Observations	29063	29063	29063	28423	29063	28176	28176	28176	27532	28176
Partial R^2	0.17	0.17	0.17	0.17	0.17	0.11	0.11	0.11	0.11	0.11
Region f.e.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Wave f.e.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Source: Author's calculations on Gallup Data. Std. errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Std. errors are clustered at the country level. OECD destinations: same as Table 3. The table shows the effect on the intention to migrate of the religious and gender egalitarian indexes aggregated with different methods: PCA (Principal Component Analysis), FA (Factor Analysis), MCA (Multiple Correspondence Analysis), KPCA (Kernel Principal Component Analysis) and AVG (same weight to each question).

Table A12: Benchmark Regressions with selected fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Var.	Int Mig	Int Mig	Int Mig	Int Mig	Int Mig	Int Mig	Int Mig	Int Mig
Destination	OECD	OECD	OECD	OECD	OECD	OECD	OECD	OECD
Religiosity	0.294***	0.304***	0.476***	0.471***				
	(0.060)	(0.061)	(0.096)	(0.102)				
Gender att.					0.096	0.198***	0.179	0.150
					(0.063)	(0.066)	(0.112)	(0.103)
Observations	30752	30752	30752	30095	29775	29775	29775	29063
Partial R^2	0.09	0.10	0.15	0.17	0.08	0.11	0.15	0.17
Country f.e.	No	No	Yes	No	No	No	Yes	No
Year f.e	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Region f.e.	No	No	No	Yes	No	No	No	Yes

Source: Author's calculations on Gallup Data. Std. errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Std. errors are clustered at the country level in columns (3),(4),(7),(8). The dependent variable is an intention to migrate dummy towards OECD destination countries. The list of OECD destination countries is the same as Table 3.

5.7.5 Matched Samples

Full Sample Matching

Table A13: Mahalanobis matching procedure 2007-2011
Matched and non-matched samples

	Non-matched Sample				*Matched Sample*			
Variables	Mean	s.d.	Diff.	P-val	Mean	s.d.	Diff.	P-val
Religiosity								
Gender	0.51	0.00	-39.49	0.00	0.45	0.01	0.82	0.68
Network	0.20	0.00	47.94	0.00	0.25	0.00	1.77	0.38
Age	33.43	0.08	-60.17	0.00	30.65	0.13	0.55	0.78
Marital	0.57	0.00	-62.44	0.00	0.47	0.01	-0.29	0.88
Children	0.74	0.00	14.05	0.00	0.72	0.00	0.84	0.67
Income	10821	89.02	7.49	0.01	11115	175.21	3.81	0.06
Education	0.47	0.00	33.52	0.00	0.52	0.01	0.75	0.71
Gender views								
Gender	0.51	0.00	-46.69	0.00	0.43	0.01	-0.50	0.80
Network	0.24	0.00	50.19	0.00	0.30	0.01	0.67	0.74
Age	33.24	0.08	-62.78	0.00	30.20	0.13	0.25	0.90
Marital	0.58	0.00	-57.22	0.00	0.49	0.01	0.92	0.65
Children	0.74	0.00	-12.37	0.00	0.73	0.00	-0.68	0.74
Income	11410	87.38	19.84	0.00	12418	173.22	4.32	0.03
Education	0.44	0.00	43.40	0.00	0.52	0.01	-0.02	0.99

The difference (Diff) between intending migrants' average covariate and intending stayers is divided by the std. error.

Arab Spring Matching

Table A14: Before- and After-Arab Spring Covariate distributions
Matched and non-matched samples

	Non-matched Sample				*Matched Sample*			
Variables	Average	Std Error	Diff.	P-value	Average	Std Error	Diff.	P-value
Gender	0.506	0.002	2.249	0.272	0.505	0.003	-2.597	0.194
Network	0.245	0.001	65.893	0.000	0.226	0.002	0.299	0.881
Age	33.660	0.044	5.676	0.005	33.739	0.071	-1.231	0.538
Marital	0.589	0.002	17.711	0.000	0.581	0.003	-4.797	0.016
Children	0.729	0.001	10.296	0.000	0.714	0.002	-1.573	0.431
Income	11577	49.101	11.885	0.000	11676	72.885	6.618	0.001
Education	0.460	0.002	-35.998	0.000	0.457	0.003	7.497	0.000

The difference individuals between average covariates of individuals before and after the Arab Spring is divided by the std error. The matching is performed on the After Arab Spring population.

Chapter 6

Network-based Connectedness and the diffusion of Cultural traits

Abstract

This paper empirically investigates the impact of network-based connectedness on the diffusion of cultural traits. Using Gallup World Poll data over 148 countries on individual connectedness, opinions and beliefs, we find that natives who have a connection abroad are associated with higher levels of social behavior, religiosity and gender-egalitarian attitudes. Due to the endogenous nature of the variables, we strongly mitigate the threat driven by selection into connectedness by showing robust estimates after controlling for broad measure of connectedness and performing propensity score and covariates matching techniques. Statistical tests are carefully implemented to quantify the selection threat driven by unobserved factors, which appears negligible. Suggesting evidence show that connectedness favors cultural convergence across regions, while increases cultural heterogeneity within regions. Exploring the mechanisms, we provide evidence that the effects are precisely estimated among less educated natives and connectedness affects economic outcomes through remittances. Differential cultural effects based on connection's residence are estimated, suggesting a destination-specific transfer of norms. Overall, the effects are sizeable at the global level on social behavior, once simulations based on estimated coefficients are performed. Although robust, gender-egalitarian and pro-religiosity effects of connectedness are limited, albeit not negligible.

6.1 Introduction

During the last decades the population of international migrants has considerably increased at the global level. The size of international migrant population reached 258 millions in 2017, while it was only 77 millions in 1960 (United Nations, 2017). Although it is still a moderate amount compared to the worldwide population (3.4% in 2017, United Nations, 2017), the role and contribution of such population changed dramatically due to recent technological development. New means of communication and globalization, in general, facilitate and increase the interactions between people living geographically apart. Individuals, thanks to interactions with peers living abroad, can learn and discover novel experiences which were hardly accessible decades ago. Such an increase in connectedness and interaction with people living abroad rises a relevant set of questions: does having a reliable connection abroad makes natives *culturally* different compared to their neighbours? How, and in which direction? Does this cultural effect of connectedness (i.e. the fact of having a reliable connection abroad) influence the distribution of cultural traits at the global level?

The present paper empirically investigates these questions using a novel and exhaustive approach. Using Gallup World Polls data on 148 countries and harmonizing intra-country regional identifiers over the 2009-2012 period, our representative sample of seven hundred thousand individuals covers around 97% of the world population over 2256 regions. We focus our analysis on natives. In order to assess the cultural effect of connectedness we built proxies of individual pro-social behavior (i.e., helping strangers and being active in the society), religiosity and gender-egalitarian attitudes. Our focus on those traits is due to their relevance on individuals preferences and countries economic growth.[1] Correlations of those cultural traits with similar indicators and deep economic preferences are provided using alternative datasets. We test the effect of transfer of cultural norms by including the average level of culture in connection's country of residence. Since the quality of the tie (i.e., kinship or not) can influence the interaction with the network member, careful robustness checks are provided accounting for the relationship with the connection abroad. To minimize the potential threat driven by unobserved omitted factors, a great variety of time and geographic fixed effects and individual controls are included in our benchmark specification.

A relevant contribution of our paper is that, by exploiting three sources of heterogeneity (country, regional and individual) we are able to look at variations of the cultural effect

[1]Several papers address the socioeconomic relevance of those traits, for instance: social behavior (Tabellini, 2010; Flanagan and Levine, 2010; Falk *et al.*, 2018), religiosity (Weber, 1946; Chase, 2014; Benabou *et al.*, 2015) and gender-egalitarian views (Baxter and Kane, 1995; Duflo, 2012; Inglehart *et al.*, 2017).

of connectedness. Using information on the average regional level of each cultural trait, cultural distances and diffusion of connectedness, our analysis shed some light on whether connectedness is bringing cultural convergence across regions and traits (Rapoport *et al.*, 2018). Moreover, we explore whether connectedness is influencing cultural heterogeneity within regions (Desmet and Wacziarg, 2018). We also investigate whether the effects are not driven by confounding factors such as countries and individual openness to foreign societies and information. Moreover, since connections abroad are sources of novel and different information, their influence on individuals' culture can vary given the information set held by each individual (Della Vigna and Gentzkow, 2010; Moriconi *et al.*, 2018). Economic channels, differential effects driven by connection's country of residence and changes in the cultural effect across separated groups of individuals are explored to clarify potential mechanisms.

Assessing the impact of connectedness on individual culture is a serious empirical challenge. The ideal comparison for our empirical question would be between two individuals which are identical, a part from experiencing an exogenous cultural shock due to connectedness (e.g. a friend which decides to move and live abroad). However, connectedness and cultural traits are endogenous and related to several individual and contextual characteristics. Due to the structure of the data (i.e. repeated cross-section) and the individual nature of our analysis, it is hard to find a source of exogenous variation which is not related to cultural traits. Nevertheless, to go beyond partial correlations and to strongly mitigate endogeneity issues driven by omitted variables and selection into connectedness, we control for more broad measures of connectedness and we combine statistical tests on selection driven by unobserved factors and matching techniques. As far as the latter are concerned, we use two different matching approaches used in the literature. The first one is a propensity score matching approach (Dehejia and Wahba, 2002; DiPrete and Gangl, 2004; Caliendo and Kopeining, 2008), which matches connected and not connected individuals with virtually the same probability of having a connection abroad. Matching methods are widely applied in nonexperimental causal studies where selection bias could arise. We estimate the probability of having a connection abroad with different models and we use a full battery of different matching algorithms to assess the robustness of the estimated effects. The second one is covariate matching approach (Imbens and Rubin, 2012; Ruyssen and Salomone, 2018; Docquier *et al.*, 2019b), which matches individuals after minimizing distances across observed characteristics. Concerning selection driven by unobserved factors, Oster (2019) provides statistical tests to evaluate the size of the bias. Moreover, Rosenbaum bounds are estimated to evaluate the robustness of matching results to unobservables (Aakvik, 2001; Rosenbaum, 2002). Overall, those methods strongly mitigate the potential selection threat and provides consistent estimates.

We show four main findings. First, we find a strong and positive correlation between having a connection abroad and each cultural trait. Individuals with a reliable connection abroad are more active in the society, as indicated by a higher level of pro-social behavior, more religious and share more gender-egalitarian views compared to the rest of the population. Second, the cultural effect of connectedness is bringing cultural convergence across regions, but only on religiosity and gender-egalitarian attitudes. The relation is stronger among individuals living in regions characterized by low levels of religiosity and gender-egalitarian attitudes, suggesting that the positive cultural effect is enhanced when the respective average cultural norm is low. Moreover, we provide evidence suggesting that connectedness is enhancing cultural diversity within regions on each trait. Third, we provide some clear insights into the channels through which connectedness affects individual culture. One significant result is that cultural changes due to connectedness are precisely estimated among less-educated rather than highly-educated individuals. This is consistent with the idea that connections abroad bring novel experiences and information to individuals, and individuals with smaller information set/previous knowledge respond more to the new source of information. Moreover, connectedness positively influence natives economic conditions through remittances. Receiving economic help from connections abroad can affects natives attitudes towards different cultural sets compared to the local ones. We also show differential effects of connectedness on connected individuals based on connections' location. Those results are suggestive evidence of a transmission of destination-specific norms and traits to natives. Fourth, we use our findings to simulate and evaluate the magnitude of the cultural effect at the individual and global level. At individual level, having a connection abroad influence natives culture in a significant way, particularly on social behaviors and gender-egalitarian views. At the global level, the cultural effect is sizeable on social behavior, while a minor effect, albeit not negligible, is predicted on gender-egalitarian attitudes and religiosity. The predicted relative distance due to connectedness, compared to a benchmark scenario where people have no connection abroad, is around of 8% for social behavior, 0.8% for gender-egalitarian views and only 0.2% for religiosity.

This paper is related and contributes to fours strands of the literature. First, this study is related with growing literature linking culture and economic development (see Guiso *et al.*, 2006, for an overview of the relation between culture and economics). Cultural aspects affect the quality of economic interactions and institutions development (Guiso *et al.*, 2009; Tabellini, 2010), the utilization of resources (Duflo, 2012) and individual preferences (Campante and Yanagizawa-Drott, 2015, Atkin, 2016). Cultural distances between countries can influence the speed of democracy transition (Murtin and Wacziarg, 2014) and knowledge diffusion (Spolaore and Wacziarg, 2012). Moreover the rising cultural diversity driven by

the increasing diversity within countries population due to migration is positively associated with economic growth (Alesina et al., 2016; Docquier *et al.*, 2019b).

Second, our paper contributes to the literature on cultural traits measurements. In the economic literature, culture is defined as a set of beliefs, attitudes and behaviors that influence individual preferences and national institutions, and remains fairly unchanged across generations (Guiso *et al.*, 2006). Trust towards other individuals, believing in the existence of an afterlife and degree of patience are all examples of cultural traits which have implications for countries and individual economic and human development (Knack and Keefer, 1997; Fukuyama, 2001; Ager and Ciccone, 2017; Falk *et al.*, 2018). Moreover, two main approaches are used when cultural traits are analyzed and measured: either having a broad approach, analyzing several aspects and proxies of culture simultaneously (see Spolaore and Wacziarg, 2015; Desmet *et al.*, 2017; Desmet and Wacziarg, 2018, and Rapoport *et al.*, 2018) or focusing on specific cultural traits, highlighting their effect on individual and national outcomes (see Tabellini, 2010; Chase, 2014 and Docquier *et al.*, 2019). We follow the latter approach by focusing on specific cultural traits, building proxies of individual social behavior, religiosity and gender-egalitarian attitudes. Our proxy of social behavior is a combination of civic engagement and inter-personal trust, which have a key role in explaining personal growth and democratic engagement, institutions and economic exchanges (Flanagan and Levine, 2010; Guiso *et al.*, 2009; Tabellini, 2010). Religiosity is negatively associated with openness to innovation and economic development (Benabou *et al.*, 2015; Chase, 2014), although is positively associated with individual well-being (Campante and Yanagizawa-Drott, 2015). Concerning gender-egalitarian attitudes, they have a pivotal role to explain actual gender-discrimination and inequalities in several spheres of the society, increasing barriers to countries development (Baxter and Kane, 1995; Duflo, 2012; Inglehart *et al.*, 2017). Using microdata on the above-mentioned cultural traits, we build novel proxies of those traits all over the world.

Third, our paper is linked with the growing literature which analyzes the determinants of cultural change and evolution. Exogenous factors like climate instability have been shown to play key roles in explaining persistence of cultural traditions and the degree of individual loss aversion (Galor and Savitskiy, 2018; Giuliano and Nunn, 2019; Sinding Bentzen, 2019). Institutions, like regulatory institutions and quality of government, have an influence on individual culture, although no causal relationship is clearly established (Alesina and Giuliano, 2015). Cultural changes are also associated to the nature of cultural transmission across generations. Using inter-generational transmission models, the persistence and the eventual evolution of culture is affected by parental preferences and costs of departure from

the dominant cultural norm (Bisin and Verdier, 2000; Baudin, 2010; Chabé-Ferret, 2019). Moreover the speed of cultural evolution changes across generations and it varies across different cultural traits (Giavazzi *et al.*, 2019). More related to our analysis, researchers have also investigated the role of modernization and globalization on cultural changes. Economic development and access to new sources of information have been analyzed as potential determinants of the rise or decline of cultural homogeneity world wide (Putnam, 2000; Inglehart and Baker, 2000; Inglehart, 2018). Desmet and Wacziarg (2018), using the General Social Survey in the US, show an increase of cultural heterogeneity from the 1990s, but it is related only to some cultural traits and social groups in the society. Access to information and interaction with individuals belonging to different social groups are underlined as leading factors of cultural evolution. Our article, using different data and approach, investigates similar determinants of cultural change through the connection and interaction with peers abroad. With a particular focus on international migration, Rapoport *et al.* (2018) provide three mechanisms through which international migrants could affect cultural convergence/divergence across countries: emigrants cultural selection, social mixing in the destination countries and social remittances from destination to origin countries. To the best of our knowledge, we are the first to provide evidence of the latter mechanism using microdata at the global level, providing an extensive analysis of the heterogeneity of the effects across regions and individuals, and also simulating the overall impact.

Fourth, our paper contributes to the literature on the role of peers and networks, in general, and migrant networks, in particular. Networks have been highlighted as conveyors of new information and behaviors. Thanks to the increased ability to communicate over distances, the flow and quality of information from personal connections plays an increasing key role in explaining individual preferences and behaviors (Granovetter, 2005; Jackson, 2014; Bailey *et al.*, 2018). Recent works have produced empirical results of the effect of the diaspora abroad on political preferences (Spilimbergo, 2009; Batista and Vicente, 2011; Docquier et al., 2016; Barsbai et al., 2017), on fertility behaviors (Beine et al., 2013) and on technological norms (Valette, 2018) of populations left-behind. However no evidence is provided on the effect on cultural traits, like gender-egalitarian attitudes, religiosity and social behavior.[2] This paper provides novel evidence of the effect of networks on individual cultural traits and behaviors.

The rest of the paper is organized as follows. Section 6.2 describes the data, the proxy of connectedness and culture traits and the correlations between those measures and alternative data sources. Section 6.3 discusses the empirical specification and the econometric

[2]Evidence is provided on the effect of returning migrants on gender-egalitarian attitudes (Tuccio and Wahba, 2015), on fertility behaviors (Bertoli and Marchetta, 2015) and on political norms (Chauvet and Mercier, 2014; Tuccio *et al.*, 2019).

challenges of our analysis. Section 6.4 shows the results of our analysis. Section 6.5 investigates whether connectedness influence cultural heterogeneity across and within regions and explores potential mechanisms. Section 6.6 presents the simulations based on our estimates. Section 6.7 concludes the paper.

6.2 Data and Summary Statistics

Individuals' culture is influenced and determined by several individual and contextual factors. The place of birth, parents' traditions and background, the education path and working experiences are all examples of aspects that influence people beliefs. In this context, interactions with other individuals (e.g. parents, friends, colleagues, etc.), with their experiences and cultural beliefs, are among the main drivers of the formation of individual cultural identity. The focus of this paper is a specific type of interaction, namely the interaction with individuals who are living in a foreign country. Those connections abroad have a unique effect on the cultural attitudes of natives. First, by sharing their experience of living in a different country, they affect individuals openness to different cultural beliefs compared to the ones shared in the local context; second, by facing foreign country's culture, those connection abroad can influence the culture of individuals in the origin countries by sharing their judgment and appreciation of the foreign culture. However, the direction of the influence on natives' culture can depend on several individual aspects associated with both the connection abroad and natives.

To test whether having connections abroad makes individuals culturally different compared to individuals who are not connected, we need micro level data on cultural traits, connectedness and other individual characteristics. Those data are available in the representative Gallup World Polls (GWP). Started in 2005, the GWP is a world-spread survey which covers over 160 countries over different aspects of individuals life, from sociodemographic characteristics to attitudes and beliefs. Our sample includes around 157 countries where Gallup conducted at least one wave of its survey over the period 2007-2016.[3] For each year and country, the sample includes around 1000 randomly selected respondents which are representative of the population aged 15 and over. For big countries, like China and Russia, the sample of respondents varies between 2000 and 5000 respondents per wave. The surveys are conducted by telephone in countries where the telephone coverage is at least 80%, otherwise face-to-face interviews are conducted in randomly selected households. Our full sample cov-

[3]The list of countries in analysis is available in Table A4 in the Online Appendix

ers around 712000 respondents aged between 15 to 90.[4] We then remove migrants from our sample, to focus our analysis on natives.

In addition to information on respondents' beliefs and connections over the majority of countries in the world, GWP includes unharmonized intra-country regional location of respondents. The combination of a wide coverage of the world and fine geographical location of respondents makes GWP an exceptional database for our research question.[5] To properly exploit the intra-country geographical location, we harmonized the GWP data to match the Database of Global Administrative Areas (GADM), which is a high-resolution database of country administrative areas on several administrative levels (from country to province level). We match the data at the regional level, when such information is available.[6] For example, such level of resolution corresponds to state-level for the US and to NUTS2-level for the majority of European Countries. The matching between GWP and GADM intra-country identifiers is perfect for the majority of individuals in our sample (93%). However, the intra-country regional location available in GWP does not correspond to a precise GADM-administrative area for the remaining 7% of respondents, but to broader geographical/administrative clusters. We then randomly distribute them to the first finer regional unit available.[7] Our harmonized sample covers 2256 regions over 148 countries and around the 97% of the world population.

6.2.1 Measuring Connectedness

The GWP question used to measure whether individuals are connected with peers abroad is the following: "Do you have relatives or friends who are living in another country whom you can count on to help you when you need them or not?". We define connected people as those who answered to this question affirmatively. The question provides not only a mere figure on whether respondents know someone abroad, but whether those connections are strong and reliable in case of need. GWP does not provide information on the characteristics or the economic status of the connection, however it provides a proxy of the degree of the relation

[4]Descriptive statistics are available in Table A5 in the Online Appendix

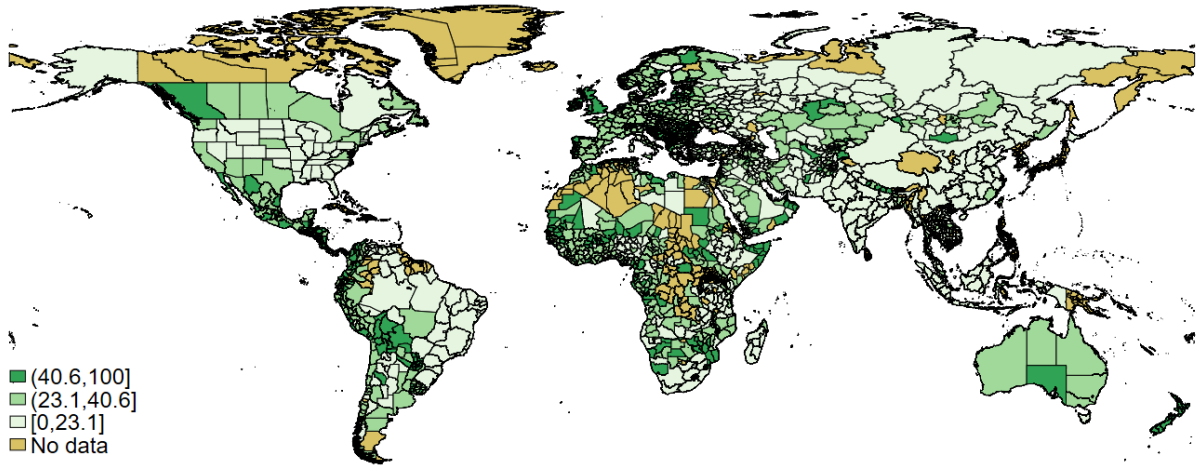
[5]Few other papers use GWP data on migration research, both at micro (Dustmann and Okatenko (2014), Bertoli and Ruyssen (2018), Ruyssen and Salomone (2018) and Docquier *et al.* (2019)), and macro level (Docquier *et al.* (2014), Docquier *et al.* (2015) and Dao *et al.* (2018))

[6]Few countries, like Croatia, Iceland, Kosovo, Luxembourg, Macedonia, Malta, Qatar, Philippines and Singapore, where impossible to match, due to the too fine geographical location of respondents in the GWP. To avoid comparison between different geographical and administrative units, we remove them from the analysis.

[7]We keep track of those respondents and broad regions through the analysis. In particular Table 6.4 provides robustness checks of our main results after dropping those randomly assigned individuals or performing the analysis with the broader geographical/administrative clusters. The main results remain unchanged.

between the respondent and the connection. A recent paper by Bertoli and Ruyssen (2018), investigating the role of networks of migrants abroad on natives preferences, shows that who answers affirmatively to this question is not only more prone to emigrate, but would like to emigrate to the country where the connection lives. Indeed, over a shorter period (2007-2012) the GWP also provides the connections' country of residence.[8] In our sample around 31% of the population has a reliable connection abroad. Although the size of the connected population is quite high compared to the share of international migrants over the world population (around 3.4 %), it is important to recall that one individual in a foreign country could be the connection for several peers at the origin country, and foreign natives could be counted as reliable connection abroad. Moreover there is large variation across countries. Countries exhibiting the highest share of connected individuals are New Zealand (73%), Ireland (73%) and Jamaica (68%), while among the countries with the lowest share of connected people there are Vietnam (7%), India (5%) and China (3%).

Figure 6.1: Geographical distribution of Connected People



Note: authors' calculations on Gallup World Poll Data. The figure plots the regional average percentage of people with a reliable connection abroad.

Figure 6.1 exploits even more the heterogeneous distribution of connected individuals by plotting the share of connected individuals at regional level. While European regions present a similar distribution of connected people, American and African regions are deeply heterogeneous. Moreover Asian regions (particularly the ones in Russia, China and India) exhibit the lowest share of connected people. The within-country variation of the share of connected people across administrative units can be even more relevant than across-country variation. For instance, in Mexico the share of connected people in the state of Quintana

[8]GWP provides up to three countries of residence, but we focus our analysis on the first answer.

Roo is 23.1%, while it reaches 65% in the Chihuahua state. The country that exhibits the highest intra-country variation is Afghanistan, where 80% of the Daykundi province population has a connection abroad, while around 2% of the population of Kapisa or Panjshir provinces are connected. Also developed countries presents large intra-country variations. In the US, almost 70% of the population in District of Columbia has a reliable connection abroad, while almost no one has it in Wyoming. Countries characterized by the lowest level of intra-country variation (maximum around 7% between the highest and lowest share of connected people) are small ones like Bahrain and Cyprus, or north European countries like Finland.

Focusing on the reported country of residence of connections, around 69% of the connections resides in an OECD high income country, while only the 31% in developing countries. The country which hosts the highest percentage of connections (i.e. people who are considered reliable connections by individuals from origin countries) is the US, which hosts around 20% of the reliable connection abroad. Other western and developed societies host a sizable percentage of connections, like Germany (7.1%), United Kingdom (6.5%) and France (6.3%). Among the non OECD high income countries, Russia (5.01%) Saudi Arabia (2.3%) and Argentina (2.01%) are the most frequently reported country of residence of reliable connections. Those figures shows an high presence of reliable connections in developed high-income societies. This distribution remains true after exploiting the heterogeneity across countries of connected people. From OECD high income countries, 86% of connected people have a connection in an OECD high-income country, while 64% of connected people in not OECD high-income country have a connection in an OECD high-income country.

6.2.2 Measuring Cultural Traits

Measuring cultural traits is not an easy task, since the definition of culture is broad and includes behaviors, way of thinking, customs and beliefs (Shenkar, 2012). According to the economic literature and Guiso *et al.* (2006), culture can be as those values and beliefs that remain fairly unchanged across generations. Moreover there are different approaches in the literature on culture and cultural aspects. On one hand, some authors do not focus on specific cultural traits. Rather they look at the whole set of related questions available in the survey in analysis (Desmet *et al.*, 2017 and Desmet and Wacziarg, 2018). On the other hand, another part of the literature prefer to focus on specific traits which are recognized as relevant for the economic development, like trust (Knack and Keefer, 1997 and Tabellini, 2010) or religiosity (Benabou *et al.*, 2015). We decide to follow the latter approach, by

focusing on three distinctive cultural traits highlighted by the literature and identifiable in the GWP: *Social Behavior*, *Religiosity* and *Gender-Egalitarian views*.

Why focusing on those specific cultural traits? The reason resides in the relevant and distinctive socioeconomic effect associated to each of them. Although through different mechanisms, those cultural traits influence deeply several socioeconomic outcomes and countries economic growth. Social-behaviors are associated with individual altruism and civic engagement, which is positive for countries democratic functioning, trust and personal growth (Tabellini, 2010; Flanagan and Levine, 2010; Falk *et al.*, 2018). Moreover those behaviors signal also a higher level of inter-personal trust and can be a one of the product of what Fukuyama (2001) defines as social capital: an informal norm that promotes co-operation between individuals, therefore economic growth.[9] The importance of religion and religiosity on shaping individual preferences and behavior is well-studied in the literature from the seminal theory of Protestan work ethic (Weber, 1946) to more recent studies on the evolution and distribution of religiosity and religious practices (Inglehart and Baker, 2000). Although there is a consensus of the effect of religiosity on several socioeconomic outcomes (e.g. fertility), the overall effect of religiosity on societies is still unclear: some papers find a negative association with individual openness to innovation (Benabou *et al.*, 2015) and economic growth (Chase, 2014), while others show an increase of individual subjective well-being due to religious practices (Campante and Yanagizawa-Drott, 2015). Finally, gender-egalitarian attitudes are relevant due to their direct impact on gender-discrimination (Baxter and Kane, 1995) and on women empowerment and economic growth (Duflo, 2012). Inglehart *et al.* (2017), using World Values Survey data, show a strong positive correlation between countries gender-egalitarian values and women empowerment.[10]

To identify an indicator of social behavior we focus on three questions available in the GWP, where the respondent has to answer[11] whether he/she have done one of the following activity during the last month:

$SB1$ How about donated money to a charity?
$SB2$ How about volunteered your time to an organization?
$SB3$ How about helped a stranger or someone you didn't know who needed help?

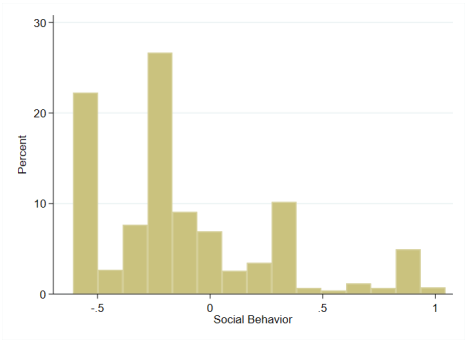
These questions are asked in all the countries and waves available in the GWP. They capture whether the respondent is actively involved in the society, helping others and interacting

[9] In the same vein, Coleman (1990) points out that social capital is embodied in the relations among persons, and it is created when the relations among persons facilitate actions.

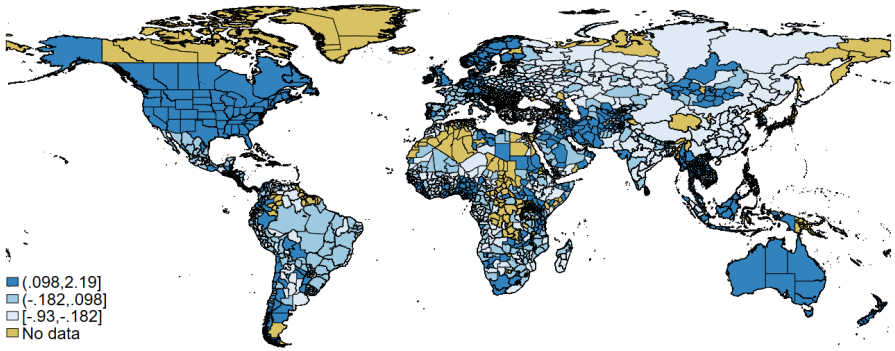
[10] The cross-country correlation between UN Gender Empowerment index and the Individual-choice index, which also captures gender-egalitarian views, is about 0.87.

[11] The respondent could answer either yes or no.

Figure 6.2: Distribution of Cultural Traits

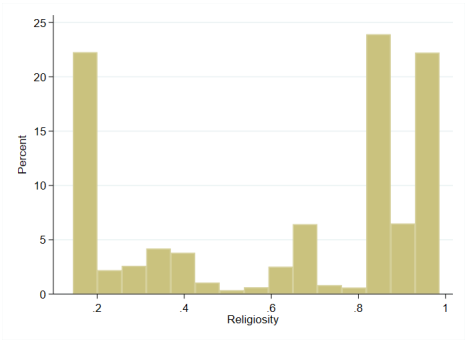


(a) Density

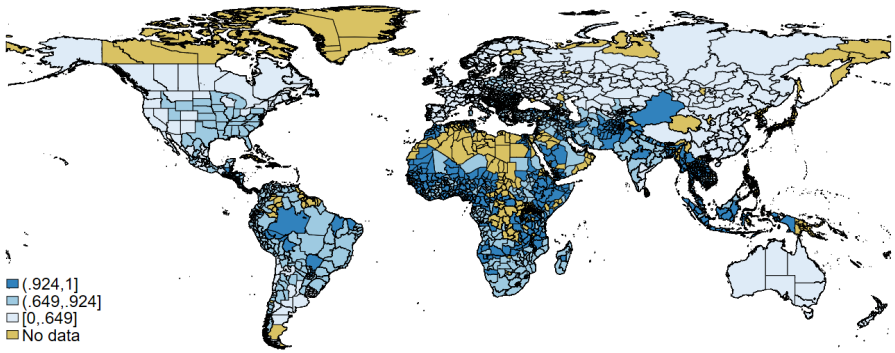


(b) Geographical Distribution

(I) Social Behavior

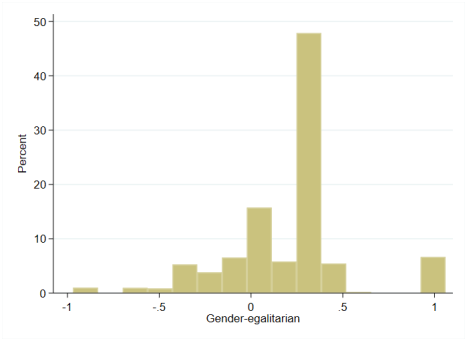


(a) Density

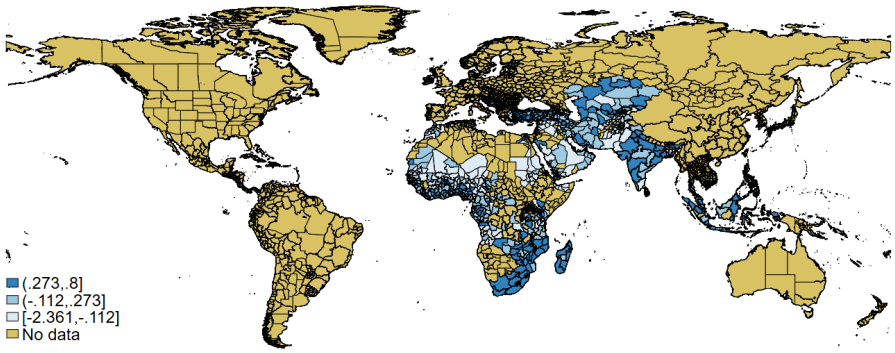


(b) Geographical Distribution

(II) Religiosity



(a) Density



(b) Geographical Distribution

(III) Gender-Egalitarian

Note: authors' calculations on Gallup World Poll Data. The Figures plots the weighted density and geographical distribution of each cultural trait: (I) Social Behavior, (II) Religiosity and (III) Gender-egalitarian views. The regions are splitted by terciles.

with other individuals in the society. We combine those three questions in one index[12] through Multiple Correspondence Analysis (MCA), which is a method to reduce the data dimensionality similar to the Principal Component Analysis, but particularly well-suited for categorical/binomial variables. Finally, we normalize the index with mean zero and standard deviation equal to one. Figure 6.2(Ia) shows the worldwide distribution of social behavior, by weighting country averages by their population. The distribution is slightly right-skewed, given the low level of social behavior and the high weight in terms of population of China. Figure 6.2(Ib) shows the geographical distribution of our index. Regions belonging to developed societies shows a distinctive higher level of social behavior compared to regions belonging to developing societies. This result should not be surprising, due to the higher availability of volunteering organizations and charity activities in western developed societies. However, a high level of heterogeneity is present among regions in developing countries, particularly in Africa and Latin America. Finally, Chinese regions exhibit the lowest level[13] of social behavior index.

The second trait in analysis is religiosity. To measure individual position toward religion, we focus on the following question:

$RE1$ Is religion an important part of your daily life?

This question is asked in almost all the countries available in the sample, excluding Jordan and Oman. Respondents who answer affirmatively to this question are defined as religious people. Figure 6.2(IIa) shows the world wide distribution of the share of religious people. The distribution looks similar to a bimodal distribution: countries are mainly characterized either by high or low share of religious people. Figure 6.2(IIb) presents the world geographical distribution of religious people. Regions belonging to Sub-Saharan African countries and Indonesia exhibit the highest share of religious people, while Chinese regions are characterized by the lowest share of religious people. Moreover western developed societies are characterized by a lower share of religious people compared to the rest of the world.

As a measure of gender-egalitarian views we follow Docquier *et al.* (2019) and we focus on three questions where the respondent has to answer whether he/she agrees or disagrees with the following statement:

$GE1$ Women and men should have equal legal rights?

[12] Correlations between questions and the synthetic index are available in Table A6 in the Online Appendix.

[13] Greif and Tabellini (2010) show that due to historical institutions and culture, China, based on Confucian moral obligation among kins and clans structure, is characterized by lower level of trust and interpersonal interactions compared to Europe, where the presence of the Church and the development of cities brought an increase of cooperation among large populations.

$GE2$ Women should be allowed to hold any job for which they are qualified outside the home?

$GE3$ Women should have the right to initiate a divorce?

We code the responses with a dummy such that having gender-egalitarian views are coded as one. Those questions are not surveyed in for the whole set of countries available in the GWP, but only to a subset of countries where gender-egalitarian attitudes are particularly salient (i.e. developing countries) and on a reduced time span (until 2011). We combine those three questions in one synthetic index of gender-egalitarian views through a Multiple Correspondence Analysis and we normalize it with mean zero and standard deviation equal to one. Figures 6.2(IIIa) and 6.2(IIIb) show the weighted density and the geographical distribution of average gender-egalitarian views among the available countries in the sample. Regions belonging to the southern Africa and Turkey are the most gender-egalitarian while Sub-Saharan regions are characterized by high gender-unequal attitudes. Afghanistan is undoubtedly the country with the lowest level of gender-egalitarian views. However, two caveats should be kept in mind. First, it is important to recall that the sample of countries includes the bottom of the world in terms of gender-egalitarian views: western and more gender-egalitarian societies are not included in the sample. Second, measures of gender-egalitarian attitudes are hardly comparable across countries, due to different contextual factors (McHugh and Frieze, 1997; Constantin and Voicu, 2015). Nevertheless, our empirical strategy will overcome such issue by comparing individuals within the same region.

Correlates of cultural traits

Those three cultural traits (social behavior, religiosity and gender-egalitarian views) are the focus of this paper. GWP is indeed a nice data source for those cultural traits, due to the geographical coverage and the availability of questions related to connections abroad. However, it is not the only data set available which encompasses those cultural aspects. Over a smaller sample of countries, World Values Survey (WVS) inquires into similar issues. Using the sixth wave of the WVS to compute country averages of similar cultural traits, we can test whether the cultural traits in analysis are indeed correlated with other similar data sources. Related to inter-personal trust and social behavior, the WVS asks: (i) "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?". Concerning religiosity, four questions from WVS are particularly relevant: (i) "How important in life is religion?"; (ii) "How often do you attend religious service?"; (iii) "How often do you pray?" and (iv) "Are you a religious person?". Concerning gender-egalitarian attitudes, WVS submits to the respondents' judgment three statements

on gender-egalitarian views: (i) "When jobs are scarce, men should have more right to a job than women"; (ii)" On the whole, men make better political leaders than women do"; and (iii) "On the whole, men make better business executives than women do". We code and normalize those questions[14] to have the same order as the GWP and also an overall index of gender-egalitarian views is created from the three above-mentioned questions.[15] Table 6.1 shows the correlation of the country averages of each cultural trait from the GWP with the related trait measured with the WVS. The maximum number of countries available in the sixth wave of WVS is 60, and only 25 of them overlap our sample of countries that covers issues on gender-egalitarian views. The reported correlations are all positive and statistically significant. The GWP measure of religiosity and the index of gender-egalitarian attitudes are highly correlated with the WVS responses. The index of social behavior is correlated with the measure of trust from the WVS, however the size of the correlation is smaller, since our index of social behavior does not capture just inter-personal trust but also active involvement in the society. Overall, those results show reassuring correlations across different data sources.

Our focus on those cultural traits is driven by their relevance to shape societies, economic development and personal well-being (Flanagan and Levine, 2010, Tabellini, 2010, Duflo, 2012, Benabou *et al.*, 2015, and Inglehart, 2018). However, to test their economic relevance, we investigate whether those cultural traits are related with deep individual economic preferences. Data on deep economic preferences are available for the 2012 wave of GWP, thanks to the Global Preferences Survey (GPS) developed by Armin Falk and coauthors. Through the framework of 2012 wave of GWP, Falk *et al.* (2018) provide six measures of individual economic preferences: patience, risk-taking, positive reciprocity, negative reciprocity, altruism and trust.[16] From their cross-country analysis, they show that patience is highly correlated with economic development, risk taking is correlated with TFP and scientific articles per capita, while negative reciprocity is correlated with the probability of armed conflicts. We regress those six measures of preferences on social behavior and religiosity, to test the re-

[14]Those questions are coded in the sixth wave of the WVS with the following codes: V24, V9, V25, V146, V147, V45, V51 and V53,

[15]All the answer are coded as dummy, which takes values of one when respondent trust others, is religious and share gender-egalitarian views. The three questions related to gender-egalitarian attitudes are combined using a MCA.

[16]Almost all the preferences are measured through a combination of qualitative and quantitative items, to validate respondent answers to the survey with behavioral experiments. "Patience" measures the propensity of the respondent to give up on something today to gain more in the future. "Risk-taking" measures respondent's propensity of taking a risk, through also lottery experiments. "Positive reciprocity" indicates the propensity to thank a stranger that help them and to return favors. "Negative reciprocity" indicates the willingness of revenge after being unjustly treated. "Altruism" quantifies how much respondent would act for a good cause without expecting anything in return. How much other people as good intentions is measured by "trust".

Table 6.1: Cultural traits Correlations - GWP and WVS

WVS Questions	(1) Social Behavior	(2) Religiosity	(3) Gender-Egalitarian	(4) Countries
Trust People (V24)	0.292**	-	-	60
Religion Important (V9)	-	0.948***	-	60
Active Rel. Participation (V25)	-	0.498***	-	60
Active Prayer (V146)	-	0.857***	-	60
Religious Person (V147)	-	0.787***	-	60
Women have Job (V45)	-	-	0.702***	25
Women be politicians (V51)	-	-	0.782***	25
Women able run business (V53)	-	-	0.745***	25
Gender-Egalitarian Index (WVS)	-	-	0.795***	25

Note: authors' calculations on Gallup World poll data and the sixth wave of the World Values Survey. The significance of the correlations are presented as follow: * p<0.1, ** p<0.05, *** p<0.01. Each column show the correlation with a specific trait computed from the GWP: social behavior (col. 1), religiosity (col. 2) and gender-egalitarian views (col. 3). Column 4 shows the number of countries.

lation between cultural traits and economic preferences. Since GPS is available only for 2012, we are unable to test the relation between gender-egalitarian attitudes and economic preferences.

Table 6.2 shows the partial correlation between social behavior (panel A) and religiosity (panel B) and economic preferences. Each column has a different economic preference as dependent variable and includes a set of individual socio-demographic characteristics.[17] The estimates in the top panel show that the index of social behavior has a positive and significant relation with all the measures of economic preferences, except for negative reciprocity. On average, an increase in the index of social behavior of one standard deviation is associated with an increase of 10% standard deviation of measures of economic preferences. Estimates relative to religiosity, presented in the lower panel of Table 6.2, show that being religious is significantly associated with a lower level of patience and negative reciprocity. Moreover religious people are more altruistic. Two conclusions can be drawn from the previous results. First, the cultural traits in analysis are indeed related with deep individual economic preferences. Knowing the importance of deep individual economic preferences on individual

[17]Each regression includes dummies on gender, marital status, having children, whether the respondent is college educated and a continuous measure of age. Moreover each dependent variable is standardized with mean zero and standard deviation equal to one.

Table 6.2: Cultural traits and Economic preferences
Global Preferences Survey

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	OLS	OLS
Year	2012	2012	2012	2012	2012	2012
Econ. Pref.	Patience	Risk Taking	Pos. Recipr.	Neg. Recipr	Altruism	Trust
Panel A - Social Behavior						
Social Behavior	0.120***	0.101***	0.097***	0.004	0.179***	0.051***
	(0.021)	(0.012)	(0.017)	(0.014)	(0.021)	(0.016)
Observations	56739	56736	57029	55863	56800	55989
Countries	59	59	59	59	59	59
Adj. R-Square	0.05	0.09	0.02	0.03	0.04	0.01
Panel B - Religiosity						
Religiosity	-0.269***	0.006	0.084*	-0.121***	0.256***	-0.010
	(0.061)	(0.041)	(0.049)	(0.035)	(0.050)	(0.042)
Observations	57886	57864	58196	56947	57952	57086
Countries	60	60	60	60	60	60
Adj. R-Square	0.05	0.08	0.01	0.03	0.02	0.01
Individual Controls	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data and Global Preferences Survey. Standard errors are clustered at country level. * p<0.1, ** p<0.05, *** p<0.01. Each regression also includes an education dummy, gender dummy, marital status dummy, having children dummy and age.

behaviors, outcomes and decisions, those correlations confirm the economic relevance of focusing on those specific cultural aspects. Second, the relation between cultural traits and economic preferences is trait specific. Social behavior is positively related with all the economic preferences which produce a positive economic outcome, like being patient or being risk-taking. Religiosity is, on one hand, negative related with patience, which is recognized by Falk *et al.* (2018) as the most relevant economic preference to explain economic development. On the other hand religiosity is associated with lower level of negative reciprocity and more altruistic preferences, which are both beneficial for social cohesion and peaceful relationships.[18]

[18]Those results confirm the Campante and Yanagizawa-Drott (2015) findings, where religious practices have a negative economic effect but at the same time positive effect of individual satisfaction.

6.3 Empirical Strategy

One of the aims of our paper is to test whether individual cultural traits are affected by connectedness. In section 6.3.1 we describe the benchmark empirical specification. Due to the empirical challenges of our analysis, section 6.3.2 presents some econometric issues driven by selection and omitted variable bias, and how we tackle them.

6.3.1 Benchmark Specifications

In this section we present our empirical approach to investigate the impact of having a connection abroad on individual cultural traits. As main explanatory variable we include $Network_{i,r,t}$, a dummy variable which takes value of one if individual i in region r in country c at year t has a reliable connection abroad, and zero otherwise. Defining $Cult_{i,r,t}$ our proxy of cultural traits (social behavior, religiosity and gender-egalitarian attitudes), we describe the basic specification as follow:

$$Cult_{i,r,t} = \alpha + \beta Network_{i,r,t} + \zeta \mathbf{\Gamma}_{i,r,t} + \eta_r + \theta_{c,t} + \epsilon_{i,r,t}. \tag{6.1}$$

The dependent variable $Cult_{i,r,t}$ represents our proxy of cultural trait, which is a standardized variable with mean zero and standard deviation equal to one for the index of gender-egalitarian views and social behavior, while it is a dummy variable for religiosity. The vector $\mathbf{\Gamma}_{i,r,t}$ contains a set of individual characteristics that can influence individual cultural traits, like sociodemographic factors (gender, age, marital status, the presence of children in the household and the household size) and socioeconomic factors (level of income per household member in international dollars, education[19] level, living in an urban area and being unemployed). Time-invariant intra-country regional factors, such as local culture and institutions, are captured by regional fixed effects η_r, while time-variant country specific factors, like economic growth and export, are captured by the country-year fixed effects $\theta_{c,t}$. We cluster the error terms at the regional level. Including this wide set of individual controls should mitigate potential omitted variable bias, however they can act as bad controls if they are determined simultaneously with our measure of connectedness (see Angrist and Pischke, 2009). We will provide robustness check without including those controls.

[19]Education is measured as a dummy variable equal to one if the respondent has at least 9 years of education

Estimating equation (6.1) using a linear model[20] produces a measure of the partial correlation between being connected and culture, captured by the estimated coefficient β. However, unobserved factors not captured by our linear model and correlated with the error term $\epsilon_{i,r,t}$ can bias the estimated effect. In particular, the lack of information related to connections' characteristics (e.g. his/her education, age, etc.) could be a relevant threat to the estimation of the true effect. For instance, if connections are culturally selected, then our model will capture the effect of being connected with individuals with a distinctive set of cultural traits rather than the effect of being connected with individuals that are experiencing different cultural norms by living abroad. Docquier *et al.* (2019) show that intending migrants from the MENA region are culturally selected: individuals who are less religious and more gender-egalitarian are more prone to emigrate to western high-income societies.[21] However, such cultural selection has a limited effect on the cultural distance between countries. To mitigate the potential bias driven by unobserved factors, we fully exploit the set of information available from GWP and we augment our basic specification in two ways.

First, knowing connection's country of residence d over the period 2009-2012, we augment the model by including in the regression an interaction term with the average culture of connection's country of residence. If connection's residence choice is driven by cultural aspect, such term should be able to act as proxy of connection's culture and control for the potential cultural selection of connection country of residence. Moreover it should also capture the cultural influence of connection's country of residence. We can describe the augmented basic specification as follow:

$$Cult_{i,r,t} = \alpha + \beta Network_{i,r,t} + \gamma Network_{i,r,t} * \overline{Cult}_d + \zeta \mathbf{\Gamma}_{i,r,t} + \eta_r + \theta_{c,t} + \epsilon_{i,r,t}. \qquad (6.2)$$

It is important to recall that regional fixed effects (η_r) capture all the time-invariant geographical and cultural distances which could affect connections location but which are common to all individuals in the same region r.[22] The time-invariant average culture of connection's country of residence d is $\overline{Cult}_d \in \{SB, RE, GE\}$. As for equation (6.1), the term $Cult$ represents each time a different cultural trait (social behavior, religiosity and

[20]It is important to recall that indexes of social behavior and gender-egalitarian views are continuous variables, while religiosity is a dichotomous variable. This imply that for the former two traits we are using an Ordinary Least Squares (OLS), while for the latter we are using a Linear Probability Model (LPM).

[21]Similarly, Berlinschi and Harutyunyan (2018) shows the intending migrants from former soviet-union countries are more politically active, critical against current institutions and tolerant towards other culture.

[22]Regional fixed effects also capture the shared probability of individuals in the same region r to have connections in countries closer to them (geographically or culturally).

gender-egalitarian views). The average cultural trait of connection's country of residence is computed after pooling all the available years in GWP. Thanks to the global scope of GWP, we can compute the term $\overline{SB}_d$ and $\overline{RE}_d$ for almost all countries of the world.[23] However, as we described in subsection 6.2.2, the index of gender-egalitarian view is computed only on a subset of countries. To cover the full set of connection's countries of residence we firstly complement the GWP with the gender-egalitarian index computed in the sixth wave of the WVS. As shown in Table 6.1, it is highly correlated with the average gender-egalitarian views at the country level computed with the GWP data. For the remaining set of countries we impute the average level of gender-egalitarian views based on their level of development.[24] As Inglehart (2018) points out, high-income countries have a distinctive set of gender-egalitarian views compared to developing countries. Estimating the parameters β and γ of equation (6.2) produces not only a measure of partial correlation between being connected and individual culture, but it also captures the effect driven by the culture of connection's country of residence. However, through our analysis we use both specifications presented in equations (6.1) and (6.2).

Second, even though we are capturing whether individuals have a strong connection abroad, the quality of the tie could influence the interactions and the cultural effect. People may interact more and may also be more influenced by close relatives rather than friends or other peers in their network. In such case, the estimated relation between our measure of connectedness and culture could be driven mainly by connections with people sharing the same kinship. Nevertheless, Granovetter (2005) points out that networks affect individuals behavior through the quality of information, and he argues that "weak ties" are more likely to transmit unique and more relevant information rather than "strong ties". Since close relatives may belong to the same cultural circles, the degree of novelty of their acquired information can be lower than friends or acquaintances which can belong to more heterogeneous and diverse groups. Similarly, Batista *et al.* (2019) show that having a network with households who experienced international migration increases political participation of individuals in Mozambique, and the effect is stronger when it is a chatting network rather than kinship network. To test whether the quality of the tie matters, we use the following question of GWP: "Have any member of your household gone to live in a foreign country permanently or temporally in the past 5 years?". If the individual answers affirmatively,

[23] Only for the 1% of the connections we do not exactly know the average cultural trait of the country of residence. We then impute to them the average cultural trait using the same imputation method as for the gender-egalitarian views.

[24] Using the available data, we compute the country average index of gender-egalitarian views for OECD high-income countries and for the rest of the world. Then we impute it to countries without the average level of gender-egalitarian, based on whether they are OECD high-income countries or not. We impute the value of gender-egalitarian views for 14% of the connections.

GWP also asks to which country did he/she moved. We then replace our main variable of connectedness in equation (6.2) ($Network_{i,r,t}$) with $Family\ Net_{i,r,t}$ or $Network^{Cl}_{i,r,t}$, respectively. $Family\ Net_{i,r,t}$ is a dummy which takes value of one if individual i has a relative abroad in any country d, while $Network^{Cl}_{i,r,t}$ is a dummy which takes value of one if individual i has a connection abroad who is not a relative in any country d (i.e. the difference between the total and family network). Estimating the partial correlations β and γ of equation (6.2) after decomposing our measure of connectedness between kinship ties and other ties sheds some light on whether the quality of ties affects the cultural effect of connectedness.

6.3.2 Econometric Issues: dealing with selection

Our empirical approach involves methodological issues that might produce inconsistent estimates of the true relation between connectedness and culture or could affect the interpretation of the estimates. A serious threat is driven by selection among connected people. If connected people are culturally selected, then our estimated coefficients are just spurious correlations of the effect of connectedness on culture. For instance, if religious individuals are more prone to have connections abroad, then a positive relation between connectedness and religiosity should be expected. Dealing with such potential threat is hard from an econometric point of view, due to scarce availability of sources of exogenous variation which are not related to individual cultural traits. Our empirical approach tries to overcome a potential selection threat by controlling for a set of relevant observable characteristics (Angrist and Pischke, 2010). Moreover, if unobserved factors are correlated with observed factors, the inclusion of observables also can reduce the selection threat due to unobservables (Altonji *et al.*, 2010, Oster, 2019). However, to properly address cultural selection into connectedness is better to disentangle such potential selection threat into two components. First, if selection into connectedness is completely unrelated to any kind of observable individual characteristic, then our approach is unable to completely rule out such threat. Depending on the direction of the relation and of the selection our estimates could be either a lower bound of the effect or a simple correlation.[25] However, claiming an endogeneity threat which is not related to any other kind of individual characteristic is quite unreasonable, since several individual and contextual aspects can influence both cultural traits and connectedness. For instance, few papers shows that individuals culture is related to weather shocks (Giuliano

[25]Individuals can be either positively or negatively cultural selected into connectedness. However, after plotting the average cultural trait associated to groups of individuals which spent from 0 to more than 12 hours in social activities with families and friends no clear cultural pattern is perceived across different groups. Results available in Figure A3 in the online Appendix

and Nunn, 2019; Sinding Bentzen, 2019), which can influence also people relations and migration preferences (Bertoli *et al.*, 2019). If we believe that endogeneity is a threat, then it is more reasonable to assume that it affects all the variables of our analysis. Nevertheless, to minimize such potential threat we test the robustness of our estimates after controlling for a more broad measure of connectedness. Since similar factors could push individuals into having reliable connections, either locally or abroad, then controlling for a more general measure of connectedness, which captures local and abroad connections, should capture common selection into having reliable connections. Second, if selection into connectedness is driven by observable characteristics or by unobservable characteristics which are related with observables, we can properly treat it with two methodologies: (i) we address the potential threat driven by selection on unobserved factors following Oster (2019) approach; (ii) we address the potential threat driven by selection on observables with matching techniques, which allow us to compare similar individuals in a set of relevant covariates.

Concerning selection into unobservables, we test whether the threat of selection driven by unobserved factors is enough to cripple our estimations. Oster (2019) provides a methodology to measure the degree of selection on unobserved variables. Based on the seminal paper of Altonji *et al.* (2005), Oster's approach rests on the assumption that the relation between treatment (in our case, connectedness) and unobserved factors can be retrieved from the relationship between treatment and observables. Given this assumption, such approach allows to compute two relevant indicators of the bias driven by selection on unobservables in linear models. Given the amount of variation of the dependent variable that we want to explain with our model (defined by the value of R-squared: $R_{max} \in [0, 1]$), Oster's methodology firstly allows to compute the degree of selection on unobservables (δ) relative to observables for which the estimated coefficient of connectedness is equal to zero. To give an intuition behind this estimator, if $\delta = 3$, then unobserved factors should be three times as important as observed characteristics to produce a partial correlation between having a connection abroad and culture equal to zero. As suggested rule of the thumb, if $\delta > 1$ then the threat driven by selection of unobservables should be minimized.[26] Moreover, the robustness of the estimates on selection on unobservables increases with an higher estimator δ. Since a relevant parameter to compute the estimator δ is the amount of variance of the dependent variable that we want to explain (R_{max}), we follow Oster (2019) suggestion which defines $R_{max} = 1.3\tilde{R}$,

[26] A value of $\delta = 1$ implies that selection on unobservables is as important as selection on observables to produce estimates equal to 0. A value of δ close to zero implies that an insignificant selection on unobservables compare to observed covariates makes the estimated effect equal to zero. Finally, the value of δ can also be negative, given the relation between observables and unobservables. The intuition related to the estimator remains the same: if $\delta < -1$ then the threat on unobservables is minimized.

where $\tilde{R}$ is the variance explained by our fully specified model.[27] Then, Oster's methodology allow us to compute the bounding values of the treatment effect after correcting for selection on unobservables (δ) and after defining the amount of variance explained (R_{max}). Precisely, we compute the identification set of the effect of connectedness when we do not adjust for unobservables ($\delta = 0$) and when selection on unobservables is as important as selection on observables ($\delta = 1$).[28] If the bounding set of the effect of connectedness does not include the zero, then it implies that our estimates are also robust after the correction for selection on unobservables. We compute the degree of selection on unobservables (δ) and the identification set of the effect for both the cultural effect of connectedness and the interaction term (namely, β and γ in equation (6.2)).

Concerning selection driven by observable characteristics, we estimate the cultural effect of connectedness after implementing matching techniques. Matching methods are widely applied in non-experimental causal studies where selection bias could arise (e.g. Dehejia and Wahba, 2002, Sianesi, 2004, DiPrete and Gangl, 2004, Caliendo and Kopeining, 2008, Ichino *et al.*, 2008). In our case, those methods allow us to compare the culture of connected people with similar not connected people in a set of relevant characteristics. Since differences in terms of observable characteristics are minimized, the differences in cultural outcomes between those selected groups of connected and not connected people can be associated only to connectedness and not to other relevant factors. In the matching literature, individuals can be matched either on the estimated probability of receiving a treatment or directly on covarites. Those matching approaches are defined as Propensity Score Matching (PSM) and Covariate Matching (CVM).[29] To deal with the potential selection bias extensively, we applied both methods in the following way.

Following Caliendo and Kopeining (2008) guidelines, we implement PSM to compute the average cultural effect of having a connection abroad on connected individuals. In this case the interaction term between connection location and average culture at the destination is not included,[30] so we just estimate the cultural effect of connectedness, disregarding con-

[27] Explaining too much of the variance of the dependent variable can bring to an excessive correction due to unobservables, when the robustness of the results can be undermined by other factors, like measurement errors. For this reason Oster (2019) defines the proper bounds of R_{max} on a set of randomized results from top journals, and the cutoff of R_{max} should allow at least 90% of randomized results to be robust to selection on unobserved factors. The suggested cutoff is 1.3 times the estimated R-squared. We test both with the cutoff 1.3 and 2 times the estimated R-squared.

[28] Both Oster (2019) and Altonji *et al.* (2005) suggest that equal selection ($\delta = 1$) is an appropriate upper bound on δ. If the estimated δ is negative, then we correct for the selection on unobservables with $\delta = -1$.

[29] Zhao (2004) describes the main differences of the two approaches, showing through Monte Carlo experiments that those different methods do not dominate each other in term of performance.

[30] To the best of our knowledge, this method allows to compute the average treatment on the treated (ATT) but implementation issues arise when interaction terms are included.

nection location. We firstly compute the individual probability to have a connection abroad (propensity score) using a probit model and a set of relevant covariates. Since one of the assumption of PSM is the Conditional Independence Assumption (CIA), which assumes that potential outcomes between treated and controls are independent on the treatment given the propensity score, then the choice of variables to compute the propensity score is crucial. On one hand, omitting important variables in the estimation of the propensity score can increase the bias in the estimates (Heckman *et al.*, 1997), on the other hand including variables that are affected by connectedness or just too many variables undermines the CIA condition and increases the variance of the estimated effect. For this reason we estimated two propensity scores with variables that are statistically significant (Heckman *et al.*, 1998): one with a set of all relevant and statistically significant covariates (*Main model*) and the other with only a subset of truly exogenous covariates, like age and gender (*Short model*). After graphically testing whether those two models satisfy the Common Support Assumption (i.e. some of the connected and not connected individuals should have the same probability of having a connection abroad), we perform matching through different matching algorithms. As a benchmark, we use Kernel (Epanechnikov) matching, a non-parametric matching estimator which uses a weighted average of not connected individuals within the kernel bandwidth to construct a counter factual outcome. We match individuals within the same region with similar propensity score, estimated with both models (main and short model). We test the quality of the matching by computing the standardized bias for each covariate (Rosenbaum and Rubin, 1985 and Sianesi, 2004) and their distribution (Heckman *et al.*, 1998 and Aakvik, 2001) after matching. Finally, we estimate the average cultural effect of connectedness on connected people for each cultural trait with bootstrapped standard errors. As robustness, we estimate the same effect using other matching algorithms suggested by the literature[31] and estimating the effect of a fake treatment over the control group. We also perform a sensitivity analysis of our treatment effect driven by a hidden bias, i.e. biased driven by selection in connectedness driven by unobserved factors, by computing the Rosenbaum Bounds (Rosenbaum, 2002). Such bounds measure how strong selection on unobserved factors should be to undermine the estimated treatment effect after matching (Aakvik, 2001 and DiPrete and Gangl, 2004).

Since with PSM methods we estimate the average culture effect of connectedness disregarding connection location (i.e. the interaction term in equation (6.2)), we rely on CVM methods to asses the robustness of our analysis once disparities in the distribution of covariates be-

[31] Precisely, we compute the average cultural effect of connectedness on connected people with the following algorithms: Kernel (Normal and Uniform), Nearest Neighbour (one or five individuals, with replacement) and Radius matching. The kernel/radius bandwidth, both for the benchmark and the robustness, is around 0.05.

tween connected and not connected people are minimized. As Imbens and Rubin (2012) point out, large distributional gaps in covariates increase the sensitivity of estimated coefficients to minor modifications in the specification. Following Ruyssen and Salomone (2018) and Docquier *et al.* (2019), we implement a design phase before the empirical analysis to create a balanced sample of individuals in terms of covariates. We match connected with not connected individuals within the same region using the Mahalanobis Metric Matching method. Such method creates a trimmed sample with equal number of connected and not connected individuals where distances in terms of observables are minimized. Moreover, if unobserved factors are related to observable characteristics (Altonji *et al.*, 2005), minimizing distances in observables should also minimize distances on unobserved factors. As for PSM, we test the quality of the matching by computing the standardized bias for each covariate. Estimating equation (6.2) on the balanced sample after CVM methods shows the robustness of our estimates from our augmented approach, after mitigating the selection bias.

6.4 Results

Following the outline of Section 6.3, our empirical analysis is structured as follows. In section 6.4.1 we investigate the effect of connectedness on different cultural traits using linear models; we also produce separate results including connection's country of residence and we test whether the findings are robust to different specifications and to family ties. In section 6.4.2 we test whether the results are robust after controlling for a measure of local connectedness (section 6.4.2) to selection on unobservables (section 6.4.2) and observables, through matching methods (section 6.4.2).

6.4.1 Benchmark Results

Table 6.3 presents our baseline results. The reported coefficients show the relation between having a connection abroad and culture, expressed with the index of social behavior in columns (1) and (2), a dummy of religiosity in columns (3) and (4) and an index of gender-egalitarian views in columns (5) and (6). The set of estimates are the product of linear models, by OLS for columns (1), (2), (5) and (6) and by LPM for columns (3) and (4). We report the estimates using both a simple specification with just connectedness variable (col. (1), (3) and (5)) and including also the interaction term with the culture of connection country of residence (col. (2), (4), and (6)). All specifications include regional dummies, country-year fixed effects and a set of individual controls.

Table 6.3: Benchmark Results

	(1) OLS 2009-2012	(2) OLS 2009-2012	(3) LPM 2009-2012	(4) LPM 2009-2012	(5) OLS 2009-2011	(6) OLS 2009-2011
	Social Behavior		*Religiosity*		*Gender-Egalitarian*	
Network	0.205***	0.196***	0.012***	0.013***	0.054***	0.046***
	(0.006)	(0.006)	(0.002)	(0.002)	(0.011)	(0.012)
Interaction		0.100***		-0.008		0.055*
		(0.013)		(0.006)		(0.030)
Education	0.261***	0.260***	-0.034***	-0.034***	0.151***	0.151***
	(0.008)	(0.008)	(0.003)	(0.003)	(0.018)	(0.018)
Female	-0.062***	-0.062***	0.060***	0.060***	0.410***	0.410***
	(0.006)	(0.006)	(0.002)	(0.002)	(0.016)	(0.016)
Married	0.078***	0.078***	0.008***	0.008***	-0.010	-0.010
	(0.005)	(0.005)	(0.002)	(0.002)	(0.010)	(0.010)
Child	0.002	0.002	0.005***	0.005***	-0.018*	-0.018*
	(0.005)	(0.005)	(0.002)	(0.002)	(0.011)	(0.011)
Age	0.000	0.000	0.002***	0.002***	-0.003***	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Urban	0.041***	0.040***	-0.018***	-0.018***	0.048**	0.048**
	(0.008)	(0.008)	(0.003)	(0.003)	(0.019)	(0.019)
Family Size	0.007***	0.007***	0.003***	0.003***	-0.002	-0.002
	(0.002)	(0.002)	(0.000)	(0.000)	(0.003)	(0.003)
Unempl.	-0.019*	-0.019**	0.001	0.001	0.004	0.003
	(0.009)	(0.009)	(0.003)	(0.003)	(0.019)	(0.019)
Income	0.006	0.006	0.002*	0.002*	0.750	0.741
	(0.005)	(0.005)	(0.001)	(0.001)	(0.536)	(0.537)
Observations	411367	411367	391901	391901	90239	90239
Regions	2097	2097	2065	2065	736	736
Adj. R-Square	0.16	0.16	0.38	0.38	0.22	0.22
Region F.E.	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. Columns (1), (3) and (5) show the estimates from the specification presented in equation (6.1) while columns (2), (4) and (6) show the estimates from the specification presented in equation (6.2). The dependent variable is: social-behavior index (col. (1)-(2)), religiosity dummy (col. (3)-(4)) and gender-egalitarian index (col. (5)-(6)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income.

Baseline estimates suggest that, on average, connectedness makes individual culturally different. Specifically, having a connection abroad is associated with an higher level of pro-social

behavior, higher religiosity and sharing more gender-egalitarian views. The coefficients are always significant at the 1% level.[32] Moreover, the interaction with the average culture of connection country of residence has a significant effect on individual social behavior and gender-egalitarian attitudes, but not on religiosity. As the literature points out, religiosity can be a rather persistent cultural trait compared to others (Inglehart and Baker (2000)). We can evaluate the economic magnitude of connectedness on each cultural trait using the standard deviations reported in Table A5 and the mean value of each cultural trait abroad.[33] Taken at their face value, the estimates in column (2), (4) and (6) indicate that having a connection in an average country in terms of culture increases individual social behavior by 0.22(=0.196+0.100*0.275) standard deviation, increases by 0.8%(=0.013-0.008*0.543) the probability of being religious and increases gender-egalitarian views by 0.078(=0.046+0.055*0.587) standard deviation.[34] This is a sizable effect. Compared to the effect of education, which appears to be one of the main determinant of each cultural traits, connectedness has a relevant magnitude. It accounts for 85% of the effect of education on social behavior and 52% of the effect of education on gender-egalitarian views. Having a connection abroad has 23% of the effect of education on religiosity, although education reduces religiosity, while connectedness has a positive effect on it.[35] Among individual controls, education, gender, marital status and place of residence (rural/urban area) have the most significant relation with each cultural trait. Women are more religious and share more gender-egalitarian views, however they are characterized by a lower level of social behavior. Having a child or being married is associated with a higher level of religiosity. While age is not associated with social behavior, older individuals are on average more religious and share less gender-egalitarian attitudes. Finally, individuals living in urban areas are characterized by more social behavior and gender-egalitarian views, but less religiosity.

Those results suggest that connectedness is bringing an overall rise of individual religiosity. However, such positive effect on individual religiosity does not imply active participation to

[32]Correcting the significance level of our estimates due to potential Multiple Hypothesis testing using Bonferroni's correction does not affect the significance level of connectedness and the interaction term on social behavior and religiosity, while the interaction term in column (6) becomes not statistically significant. However, due to the open debate on statistical significance in the field of statistics (see Wasserstein *et al.*, 2019), those results should be taken with caution.

[33]The average culture of connections' country of residence is 0.275 for social behavior, 0.543 for religiosity and 0.587 for gender-egalitarian views.

[34]Being aware that using a linear model with a dichotomous variable can be risky due to the possible prediction of the model outside the range 0-1 and the inappropriate use of linear tests, Hellevik (2009) points out that those are minor arguments and linear models should be used, particularly for a causal analysis and for the easy interpretation of the coefficients.

[35]Evaluating the magnitude of the effect of connectedness disregarding the interaction term produces similar results: an increase of 0.20 standard deviation for social behavior, an increase of 0.054 standard deviation for gender-egalitarian views and an increase of 1.2% of the probability of being religious.

place of worships or an increase of trust towards religious organizations. Table A7 in the Online Appendix shows that connectedness is not related with active religious participation. Moreover, the pro-religiosity effect is mainly present in individuals who already belong to religious groups, like Christians and Muslims.

Table 6.3 shows that connectedness is associated with a more active participation in the society, given by an higher level of pro-social behavior, and more gender-egalitarian attitudes. Those effects can be affected by individuals' perception of the society. Unfair and discriminatory conditions can enhance the desire of an alternative and fairer scenario, which can be shared and shown by connection living in more fair societies abroad. Table A8 in the Online Appendix shows the cultural effect of connectedness once perceived justice towards poor people, migrants and women are taken into account. Perceived discrimination does not influence the pro-social behavior effect of connectedness, while it enhances the positive effect on gender-egalitarian attitudes.

Robustness Checks

This subsection investigates the robustness of the previous results. Table 6.4 summarizes the findings for connectedness on social behavior (Panel A), on religiosity (Panel B) and gender-egalitarian views (Panel C) according to the benchmark model presented in equation (6.2). Each column presents a different robustness check. In column (1) we estimate our model by avoiding the imputation of the average culture of connection country of residence, which is particularly relevant for gender-egalitarian values, since we imputed average culture for 14% of the connections. Column (2) shows the estimates when we do not include individual controls, while columns (3) and (4) do not include regional and country-by-year fixed effects but include respectively country or region-by-year fixed effects. In column (5) we estimate our model after removing all the individuals with an imprecise regional identifier, while column (6) shows the estimates using broader regions as geographical units. Finally, in columns (7) and (8) we investigate whether the effects are affected after removing from the sample regions that appears only once or after removing regions with less then one hundred observations.

Overall, we show that the estimates presented in Table 6.3 are robust to different variations of the specification and of the sample. The coefficients associated to connectedness and the interaction term are always positively related to social behavior and significant at 1% level. Also the size of the coefficients remains rather similar across different robustness checks.

We then investigate whether the quality of the tie with the connection abroad influences the

Table 6.4: Benchmark Results - Robustness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dest. Clean	No Controls	Country F.E.	Reg.-Year F.E	No Imp.	Broad Reg.	Single Reg.	Small Reg.
Panel A - Social Behavior (OLS)								
Network	0.196***	0.212***	0.199***	0.192***	0.194***	0.199***	0.197***	0.195***
	(0.006)	(0.006)	(0.009)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Interaction	0.100***	0.097***	0.087***	0.095***	0.114***	0.102***	0.099***	0.103***
	(0.013)	(0.013)	(0.016)	(0.013)	(0.015)	(0.014)	(0.014)	(0.014)
Observations	411367	473340	448138	411355	369481	417907	396898	372419
Regions/Countries	2097	2221	144	2095	1749	1075	1907	1311
Adj. R-Square	0.16	0.15	0.12	0.19	0.16	0.15	0.16	0.15
Panel B - Religiosity (LPM)								
Network	0.013***	0.007***	0.013***	0.012***	0.013***	0.014***	0.013***	0.013***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)
Interaction	-0.006	-0.005	-0.001	-0.004	-0.004	-0.008	-0.008	-0.006
	(0.006)	(0.005)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Observations	391901	451701	428132	391893	350015	398441	378628	353011
Regions/Countries	2065	2190	141	2064	1717	1045	1888	1284
Adj. R-Square	0.38	0.36	0.34	0.39	0.38	0.37	0.38	0.38
Panel C - Gender-Egalitarian (OLS)								
Network	0.051***	0.050***	0.059***	0.043***	0.046***	0.040***	0.046***	0.046***
	(0.011)	(0.011)	(0.019)	(0.011)	(0.012)	(0.014)	(0.012)	(0.012)
Interaction	0.044	0.041*	0.015	0.057*	0.051*	0.061**	0.048	0.051
	(0.028)	(0.025)	(0.051)	(0.029)	(0.031)	(0.030)	(0.030)	(0.031)
Observations	90239	113308	93148	90238	87388	90223	84776	84521
Regions/Countries	736	862	58	736	593	390	685	513
Adj. R-Square	0.22	0.18	0.17	0.24	0.22	0.21	0.22	0.22
Individual Controls	✓		✓	✓	✓	✓	✓	✓
Region F.E.	✓	✓			✓		✓	✓
Broad Region F.E.						✓		
Country F.E.			✓					
Region-year F.E.				✓				
Country-year F.E.	✓	✓			✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level, a part from columns (4)-(5), where they are clustered at country level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (Panel A), religiosity dummy (Panel B) and gender-egalitarian index (Panel C). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. Each columns estimates the benchmark model in equation (6.2) by: not imputing the average culture of country of residence of connection (col. 1), not including individual controls (col. 2), removing region and country-year fixed effects and including just country fixed effects (col. 3) or region-year fixed-effects (col. 4), removing by the sample all the individuals which we impute their region (col. 5), performing the analysis on broad regions rather then decomposing them (col. 6), performing the analysis on the subsample that appears twice (col. 7) and removing regions with less then 100 observations (col. 8).

cultural effect of connectedness. Knowing whether individuals have relatives abroad, Table 6.5 shows the estimates of our benchmark model when we split our variable of connectedness between family connection (*Family Net.*) and other connections ($Network^{Cl}$). We estimate our model by including those two types of connection separately, with their respective interaction term associated to the average culture at connection's country of residence. In columns (1), (3) and (5) we show the linear estimates using not relative connectedness, while in columns (2), (4) and (6) we present the estimates using kinship connectedness. The first finding is that, in general, the point estimates associated to connectedness after removing

parental ties remain quite consistent, both in size and significance, with our previous results: having a connection abroad makes individuals more social, religious and gender-equal. Having a family connection abroad still has a positive effect on all cultural traits, although it becomes not significant for religiosity. Possibly, family connections are bringing less novelties to the information set of individuals, in particular related to a delicate trait like religiosity. Finally, evaluating the economic magnitude of those estimates, we find rather similar results between the different specifications and our benchmark estimates presented in Table 6.3.

Table 6.5: Broad Connections or Household Connections?

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	LPM	LPM	OLS	OLS
	2009-2012	2009-2012	2009-2012	2009-2012	2009-2011	2009-2011
	Social Behavior		*Religiosity*		*Gender-Egalitarian*	
$Network^{Cl}$	0.195***		0.014***		0.043***	
	(0.006)		(0.002)		(0.012)	
$FamilyNet.$		0.163***		0.006		0.076*
		(0.019)		(0.013)		(0.040)
$Interaction$	0.082***	0.170***	-0.011*	0.011	0.052*	0.029
	(0.014)	(0.032)	(0.006)	(0.019)	(0.030)	(0.063)
Observations	403521	411367	384145	391901	87756	90239
Regions	2097	2097	2065	2065	736	736
Adj. R-Square	0.16	0.15	0.38	0.38	0.22	0.22
Individual Controls	✓	✓	✓	✓	✓	✓
Region F.E.	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (col. (1)-(2)), religiosity dummy (col. (3)-(4)) and gender-egalitarian index (col. (5)-(6)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. *Family Net.* is a dummy equal to one whether the individual has a relative abroad, while $Network^{Cl}$ is a dummy equal to one weather the individual has a connection abroad who is not a relative.

Then, our measure of connectedness could capture an overall individual taste and preference for foreign knowledge. In that case, the cultural effect of connectedness is not driven by

the interaction with the connection abroad, but captures a broader individual propensity to be exposed and assimilate extra-national norms. Moreover, country openness to foreign activities can also influence individual attitudes towards novelties coming from abroad. We test the robustness of our estimates to a broader taste and exposure to foreign experiences in two step. We firstly test whether country openness influence the cultural effect of connectedness, by estimating our benchmark equation on subsamples of regions based on country development and world exposure, measured by GDP per capita, import and export.[36] We then include in our regression model additional individual controls which capture an open attitude towards the international world or higher exposure to not local information. As proxies for openness towards foreign societies, we include whether the individual would like to move abroad permanently and whether he/she is actually planning to move abroad permanently.[37] To capture the exposure to not local information we include dummies on whether the individual has internet and land-line telephone in his/her house.[38]. The estimates of both tests are presented in Tables A9 and A10 in the Online-Appendix. Overall the estimates associated to connectedness and the interaction term remain stable and precise across different specifications.

We test whether the results presented in Table 6.3 are robust to alternative estimation methods rather then simple linear models. Following Galor and Savitskiy (2018) and using a Probit model for religiosity and Ordered Probit for social behavior and gender-egalitarian attitudes, we estimate the probability of being religious and ranked level of social behavior and gender-egalitarian attitudes conditional on connectedness. Table A11 in the Online Appendix shows that using non-linear models produces estimates in line with our benchmark approach. As additional robustness check, we investigate whether the results are not driven by the way we construct our indexes of social behavior and gender-egalitarian views. Table A12 in the Online Appendix shows the estimates of connectedness on each item of the two indexes using Linear Probability Models. Connectedness has a positive and significant effect on each cultural question. Finally, we test whether the results are driven by geographically small countries. Individuals can easily interact with peers in smaller countries, due to a higher density of individuals compared to wide countries. Table A13 in the Online Appendix shows

[36]Data on GDP per capita, import and export as share of GDP at the country level are provided by the World Bank over the period of analysis.

[37]GWP includes questions on migration intentions which are included as dummy variables. Those variables takes value of one whether the respondent answer affirmatively to the following questions, respectively: "Ideally, if you had the opportunity, would you like to move permanently to another country, or would you prefer to continue living in this country?" (Migration Int.); "Are you planning to move permanently to that country in the next 12 months, or not?" (Migration Plan.). Descriptive statistics of those variables are available in Table A5

[38]Around 37% of the population has an internet connection, while 47.8% of the population has a land-line telephone in their house.

the estimates after dropping small countries (i.e. with at most 200000 KM^2 of land area). The cultural effect of connectedness remains unchanged across samples and cultural traits.

6.4.2 Dealing with Selection

In this section we investigate the likelihood that our results are affected by selection bias. We first investigate whether our relation between connectedness and cultural traits is not rule out by an overall individual propensity to having networks and connection in section 6.4.2. We test for the threat driven by selection on unobservables in section 6.4.2, applying the methodology suggested by Oster (2019). In section 6.4.2 we test whether the results are robust after controlling for selection on observable characteristics through matching methods.

Controlling for General Connectedness

As we state previously, dealing fully with the endogeneity threat between connectedness and culture at individual level is extremely hard from an econometric point of view. Particularly, if individuals' selection into connectedness is positively associated with cultural traits and it is also completely unrelated with observable characteristics, then any causal claim associated to our estimates would be outrageous. However, if we believe that endogeneity is an issue, then it would be hard to claim that it does not involve other individual characteristics. To partially address this issue we include in our regression model a more broad measure of individual connectedness. Intuitively, if cultural factors influence individual propensity to have connections (locally or abroad), then controlling for a more broad measure of connectedness should capture for the common selection into connectedness. Controlling for connectedness (in general) should then reduce the threat driven by individual selection into connectedness (abroad).

As general measure of connectedness we use the following question from GWP: "If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?". We code such question as a dummy variable ($Rel.\ Connection$) which takes value one if individuals answer affirmatively. Individuals who answer such question in an affirmative way are then characterized by a reliable connection, either locally or abroad. The correlation between $Rel.\ Connection$ and $Network$ is around 0.148, which allow us to include both variables in the same regression. Moreover, such small correlation suggests that not all the individuals who have a reliable connection in general have also a reliable connection abroad: only the 35% of them has also a connection abroad.

Table 6.6: Controlling for General Connectedness

	(1) OLS 2009-2012	(2) OLS 2009-2012	(3) LPM 2009-2012	(4) LPM 2009-2012	(5) OLS 2009-2011	(6) OLS 2009-2011
	Social Behavior		*Religiosity*		*Gender-Egalitarian*	
Network	0.195***	0.186***	0.011***	0.011***	0.041***	0.035***
	(0.006)	(0.006)	(0.002)	(0.002)	(0.012)	(0.012)
Interaction		0.102***		-0.004		0.060
		(0.014)		(0.006)		(0.037)
Rel. Connection	0.106***	0.106***	0.020***	0.020***	0.052***	0.052***
	(0.007)	(0.007)	(0.002)	(0.002)	(0.014)	(0.014)
Observations	378177	378177	362070	362070	75811	75811
Regions	2084	2084	2045	2045	709	709
Adj. R-Square	0.17	0.17	0.38	0.38	0.23	0.23
Region F.E.	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. Columns (1), (3) and (5) show the estimates from the specification presented in equation (6.1) while columns (2), (4) and (6) show the estimates from the specification presented in equation (6.2). The dependent variable is: social-behavior index (col. (1)-(2)), religiosity dummy (col. (3)-(4)) and gender-egalitarian index (col. (5)-(6)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income.

Table 6.6 is structured as our benchmark Table 6.3 after including as additional control our measure of general connectedness. Even though our measure of general connectedness is significantly related with all the three cultural traits, the estimates associated to connectedness abroad still remains positive and statistically significant. Moreover, the general reduction of the coefficients' size suggests that part of the potential positive selection into connectedness is now captured by our variable of general connectedness *Rel. Connection*. Table A15 in the online Appendix shows similar results after performing a subsample analysis based on individuals answer to our measure of general connectedness. Overall, those results shows that the positive relation between connectedness and cultural traits is robust after controlling for the potential positive selection, which is partially captured by a more broad individual's propensity of having a trustworthy connection.

Selection on Unobservables

We now investigate whether our results can be driven by selection on unobserved factors. To test the robustness of our previous estimates to selection on unobserved characteristics, we implement the methodology recently proposed by Oster (2019) and we compute for each estimated model: (i) the degree of selection on unobservables relative to observed characteristics for which the estimated coefficient associated to connectedness is equal to zero (δ), and (ii) a bounded identified set of the effect of connectedness on culture after correcting for a reasonable degree of selection on unobservables (i.e. as important as observable characteristics). Since both the estimator and the bounding set of the effect are computed based on the amount of variance that our model would like to explain (defined as R_{max}), and given $\tilde{R}$ the R-squared of the model with all the controls, we implemented Oster's methodology with the suggested bounded value ($R_{max} = 1.3\tilde{R}$) and an over-explaining value of R_{max} ($R_{max} = 2\tilde{R}$).

Table 6.7 presents the results associated to the threat of selection on unobserved factors. Each panel shows the results for each cultural trait in analysis: social behavior (panel A), religiosity (panel B) and gender-egalitarian views (panel C). Column (1) report the estimates associated to our benchmark model, while the degree of selection on unobservables and the bounding set of the estimates with different degree of R_{max} are presented from column (2) to (5).[39] Focusing on the results associated to the suggested level of $R_{max} = 1.3\tilde{R}$ in columns (2) and (3), there are two main findings. First, our estimates are robust to selection on unobservables. The degree of selection on unobservables relative to observables (δ) is above the cut off of one[40] (i.e. when selection on unobservables should be as important as selection on observables to invalidate the estimated coefficient), excluding the interaction term associated to social behavior. Moreover, column (3) shows the bounding set of the estimates when we do not correct for unobservables (i.e. $\delta = 0$) and when we correct for a reasonable degree of selection on unobservables (i.e. $\delta = -1/1$). On average, the bounding set does not include the zero, suggesting that the estimates are statistically different from zero after correcting on selection on unobservables. Second, the estimates associated to the interaction term compared to the estimates associated to connectedness are less robust to selection on unobservables. As a matter of fact, the estimated values of δ are lower for the interaction terms. This result is not surprising: the cultural effect of connectedness can be influenced by the characteristics of the connection, like his/her education or age, and also by connection's experience in the foreign country. An higher implication and exposure to the local culture in the country of

[39]The analysis associated to the estimates of the model presented in equation (6.1) are available in Table A16 in the Online Appendix. Overall the results are in line with Table 6.7.

[40]Sometimes δ can be negative, due to the correlation between observables and unobservables is negative rather then positive.

Table 6.7: Selection on Unobservables

	(1)	(2)	(3)	(4)	(5)
		$R_{max} = 1.3\tilde{R}$		$R_{max} = 2\tilde{R}$	
	Benchmark	δ	*Id. Set*	δ	*Id. Set*
Panel A - Social Behavior (OLS)					
Network	0.196***	4.860	[0.171; 0.196]	1.446	[0.082; 0.196]
	(0.006)				
Interaction	0.100***	0.773	[-0.031; 0.100]	0.208	[-0.416; 0.100]
	(0.013)				
Adj. R-Square ($\tilde{R}$)	0.16				
Panel B - Religiosity (LPM)					
Network	0.013***	-4.563	[0.008; 0.013]	-1.275	[0.002; 0.013]
	(0.002)				
Interaction	-0.008	2.197	[-0.008; -0.005]	0.614	[-0.008; 0.030]
	(0.006)				
Adj. R-Square ($\tilde{R}$)	0.38				
Panel C - Gender-Egalitarian (OLS)					
Network	0.046***	-4.970	[0.034; 0.046]	-1.346	[0.009; 0.046]
	(0.012)				
Interaction	0.055*	-2.517	[0.030; 0.055]	-0.680	[-0.018; 0.055]
	(0.030)				
Adj. R-Square ($\tilde{R}$)	0.22				

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level in column (1). * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (Panel A), religiosity dummy (Panel B) and gender-egalitarian index (Panel C). Column (1) show the estimates of the augmented model presented in equation (6.2). Columns (2) and (4) shows the value of selection on unobservables (δ) which produces $\beta = 0$ given the value of R_{max}. Columns (3) and (5) shows the identified set of the estimated $\hat{\beta}$ when $\delta = 0$ (no bias-adjustment) and $\tilde{\beta}$ when $\delta = 1/-1$ (observables as important as unobservables) given the value of R_{max}. Columns (2) and (3) shows the results for the suggested level of R_{max} by Oster (2019).

residence could magnify/reduce the amount and quality of information communicated to peers in the origin countries, influencing henceforth the cultural effect. The lack of relevant information about connection experience and characteristics is a defect of our data source and also of our analysis.

Columns (4) and (5) presents the results after increasing the amount of dependent variable variance that we want to explain. Being aware that over-controlling for selection on unob-

servables could invalidate our estimation, the degree of selection on unobservables remains above the cut off level of one for each cultural trait, but not for the interaction terms. Overall, those results provide a reassuring answer to the threat driven by selection on unobservables.

Matching Results

This section presents the results of our analysis using matching methods. Those broadly used methods are particularly popular in non-experimental causal studies, since they are able to create a relevant control group to test the effect of a treatment on the treated population. As we explained in section 6.4.2, we implement both Propensity Score Matching (PSM) and Covariates Matching (CVM): the former to capture the cultural effect of connectedness on connected people while the latter is able to include also the information related to connection location in the estimates. We present the results associate to PSM first, and then we move to CVM.

Propensity Score Matching - We follow Caliendo and Kopeining (2008) guidelines to properly implement PSM methods. First, we compute the probability to have a connection abroad using a probit model.[41] The choice of the variables in the model is crucial: on one hand all the variables that affect both connectedness and culture should be included in the model (Heckman *et al.*, 1997); however those variables should be exogenous to having a connection abroad to avoid invalidation of the CIA. We then estimated two probit models: the first (called *Main model*), includes all the variables that are significantly related with connectedness, and that should be exogenous. The variables included are age, gender, education, marital status, rural/urban location and unemployment status.[42] The second model (called *Short model*), has a more parsimonious specification, including only purely exogenous variables like gender, age and age-transformations. Table A1 in the Appendix presents both estimated models. We use both models to predict the propensity score used for the matching. Performing matching with both models allow us to check whether the effects may be driven by the selection of the variables in the probit model. Both model satisfied the Common Support assumption, as Figure A1 in the Appendix shows. This result implies that we can find for each connected individual at least a proper counterfactual not connected individual with the same probability of having a connection abroad.

[41] In case of dicothomous variables, Caliendo and Kopeining (2008) points out that logit and probit models produce similar results

[42] Column 1 of Table A1 shows the estimates when we also include family size and children dummy. Since estimates associated to those variables are not statistically significant, we remove them from the main model. Income is not included since it can be easily an endogenous variable. We include age transformation, to increase the fit of the model.

We matched connected and not connected individuals within the same region using a Kernel Epanechnikov algorithm. As bandwidth of our kernel we set 0.05 as in DiPrete and Gangl (2004). The advantage of this approach is the utilization of all the information available to build a synthetic but close counterfactual. However, individuals that could be bad matches can be included. For this reason we also perform, as a robustness, matching with alternative algorithms.[43] We matched individuals using propensity score estimated both with the main and short model. To asses the quality of the matching we test first the covariates balance between connected (treatment group) and not connected (control group) individuals before and after the matching. We compute the standardized bias $SB(X)$ (Rosenbaum and Rubin, 1985, and Sianesi, 2004) for each covariate X as follow:

$$SB(X) = 100 * \frac{\overline{X_C} - \overline{X_{NC}}}{\sqrt{[V_C(X) + V_{NC}(X)]0.5}} \tag{6.3}$$

where $\overline{X_C}$ and $V_C(X)$ are the sample mean and variance for the connected population, while $\overline{X_{NC}}$ and $V_{NC}(X)$ are the sample mean and variance for the not connected population. The standardized bias is a convenient way to measure the distance in each observable between the two groups. We compute the average and standardized bias of each variable used in the benchmark model before and after matching individuals, using both estimated propensity scores.[44] Table A2 presents the results. Columns (1) to (3) presents the results for the whole sample, while columns (4) to (6) and (7) to (9) presents the results after matching with the propensity score from the main and short model, respectively. Each panel presents the results associated to each cultural trait: social behavior (panel A), religiosity (panel B) and gender-egalitarian views (panel C). Given as |5%| a reasonable level of standardized bias (Caliendo and Kopeining (2008)), column (6) shows that matching with propensity score from the main model reduces the biases of all the covariates below the reasonable threshold (excluding age in Panel A and B). Matching individuals with the propensity score computed with the short model is able to reduce significantly the standardized bias across the covariates, although it is not completely able to reduce education disparities (column (9)). Those results shows that, after matching, connected and not connected individuals are highly comparable in terms of observable characteristics.

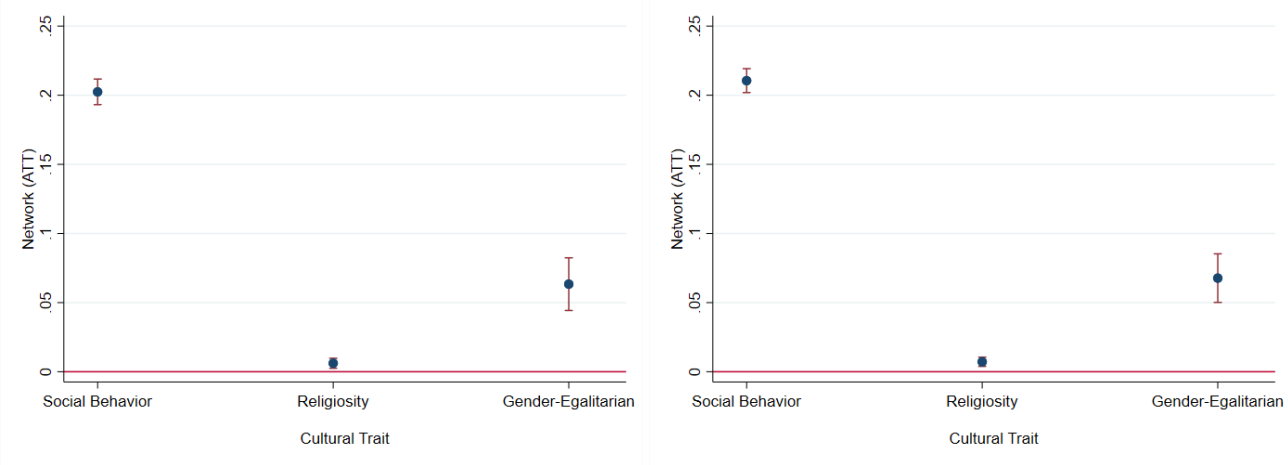
We also test whether there are differences in the distribution of covariates between connected and not connected individuals. Heckman *et al.* (1998) points out that covariates distribution

[43]The alternative matching algorithms used are: Kernel (Normal and Uniform), Nearest Neighbour (one or five individuals, with replacement) and Radius matching.

[44]The denominator of equation (6.3) remains constant before and after matching, and it is computed on the overall population.

differences between treated and controls could generate another source of selection bias. Figures A5, A6 and A7 in the Online Appendix plot the distribution of each covariate after matching with the main model between connected and not connected individuals for each cultural trait. The figures show similar observables distributions after matching, minimizing another potential source of selection bias (Aakvik, 2001).

Figure 6.3: Matching Results - Average effect of Connectedness



(a) Main Model

(b) Short Model

Note: authors' calculations on Gallup World Poll Data. The figure plots the average effect of connectedness after a propensity score matching on three different cultural traits (Social Behavior, Religiosity and Gender-Egalitarian) and the interval of confidence at 99% level. Standard errors are bootstrapped. Figure (a) shows the results using the main model presented in column (2) in Table A1 to compute the propensity score for the matching, while Figure (b) shows the results using the short model presented in column (3) in Table A1. The matching method is Kernel Epanechnikov matching.

Figure 6.3 plots the average cultural effect of connectedness on connected individuals on each cultural trait, using Kernel Epanechnikov matching and propensity score computed from the main model (a) and from the short model (b). The results confirm the estimates from our benchmark linear model: having a connection abroad is positively associated with higher social behavior, religiosity and gender-egalitarian views. Also the size of the effect is rather similar to the estimates presented in columns (1), (3) and (5) of Table 6.3: having a connection abroad is associated with an increase of social behavior of 0.197, an increase of 0.6% of the probability of being religious and an increase of gender-egalitarian views of 0.065 standard deviation. The estimates are slightly bigger after matching with the propensity score from the short model (figure b), but overall the effect are always statistically significant at 1% level. It is important to recall that those results do not account for the effect driven by the location of the connection, which are taken into account after direct matching on the covariates. Figure A4 in the Online Appendix shows that the average cultural effects

of connectedness on each cultural trait are robust after using different matching algorithms: Kernel (Normal and Uniform), Nearest Neighbour (one or five individuals, with replacement) and Radius matching.

As a robustness, we perform a second round of matching over the individuals belonging to the control group after randomly assign to individuals a fake treatment drawn from an uniform distribution. Figure A2 in the Appendix shows the average effect of the fake treatment on each cultural trait after performing the same PSM approach designed above. The fake treatment does not produce any significant cultural effect on the treated group. This result mitigate the threat driven by potential selection into the control group.

Finally, following the sensitivity analysis proposed by Rosenbaum (2002), we test the robustness of our estimated cultural effects to unobserved factors after matching methods. Matching methods manage to eliminate bias driven by selection into observable characteristics, by minimizing differences between connected and not connected individuals. However, as for linear methods, they are not robust against "hidden bias" (DiPrete and Gangl, 2004): unobserved factors that affect simultaneously individual culture and conncetedness. The methodology and the results of our sensitivity analysis are presented in 6.8.2. Overall, the results shows that the cultural effect of connectedness is strongly robust to hidden bias on social behavior and gender-egalitarian attitudes, while slightly less robust on religiosity.

Covariates Matching - An alternative way to perform matching between connected and not connected individuals is to match individuals directly on covariates instead of the probability of having a connection abroad. We matched connected and not connected individuals within each region using Mahalanobis Metric Matching method to minimize covariates distances across individuals.[45] Since such method produces a sample with equal number of connected and not connected individuals within each region, the sample size is significantly trimmed and extremely balanced. Columns (10) to (12) of Table A2 show that the quality of the matching is indeed satisfying, looking both at the average and standardized bias of each covariate.

Table 6.8 presents the linear estimates over the matched and trimmed samples of individuals. Columns (1) to (3) show the effect of connectedness on social behavior, columns (4) to (6) on religiosity and columns (7) to (9) on gender-egalitarian views. The table also reports the estimates presented in Table 6.3, to facilitate the comparison with the results on the matched sample. Overall, all the estimates of the benchmark regressions hold when using matched samples. The size of the estimates is quite similar between the benchmark results and the

[45]Such approach is suggested by Imbens and Rubin, 2012 and Ruyssen and Salomone (2018) and Docquier *et al.* (2019) provides applications of this method.

Table 6.8: Mahlanobis Matched Sample Analysis

	(1) OLS	(2) OLS	(3) OLS	(4) LPM	(5) LPM	(6) LPM	(7) OLS	(8) OLS	(9) OLS
	Social Behavior			*Religiosity*			*Gender-Egalitarian*		
Sample	Main	Matched	Matched (RW)	Main	Matched	Matched (RW)	Main	Matched	Matched (RW)
Network	0.196***	0.194***	0.190***	0.013***	0.012***	0.014***	0.046***	0.050***	0.056***
	(0.006)	(0.007)	(0.007)	(0.002)	(0.003)	(0.003)	(0.012)	(0.016)	(0.017)
Interaction	0.100***	0.097***	0.104***	-0.008	-0.001	-0.005	0.055*	0.050	0.034
	(0.013)	(0.013)	(0.015)	(0.006)	(0.006)	(0.007)	(0.030)	(0.031)	(0.035)
Observations	411367	241537	190470	391901	231198	182776	90239	49527	39762
Regions	2097	2034	2034	2065	2001	2001	736	644	644
Adj. R-Square	0.16	0.16	0.16	0.38	0.36	0.36	0.22	0.22	0.23
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region F.E.	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (col. (1)-(3)), religiosity dummy (col. (4)-(6)) and gender-egalitarian index (col. (7)-(9)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. Columns (1), (4) and (6) shows the benchmark estimates presented in Table 6.3. Columns (2), (5) and (8) shows the estimates on the matched sample using a Mahlanobis Metric Matching procedure and associated weights. Columns (3), (6) and (9) shows the estimates on the matched sample using a Mahlanobis Metric Matching procedure and reweighted weights (matching weights*survey weights).

matched results. Moreover, re-weighting the sample of individuals combining both survey weights and matching weights does not modify the size and significance of the estimates (col. (3),(6),(9)). After removing unbalanced distribution of observable characteristics through covariate matching, connectedness is still associated with an higher level of social behavior, religiosity and gender-egalitarian views.

6.5 Cultural convergence/divergence and mechanisms

After identifying a positive and significant effect of connectedness on each individual cultural trait, in this section we increase the understanding of our results, by exploring whether connectedness is bringing cultural convergence or divergence across and within regions (section 6.5.1) and potential channel of transmission of the cultural effect, through spill-over effects, economic channels and destination-specific cultural effects (section 6.5.2).

6.5.1 Cultural Convergence or Divergence

Our analysis show a strong and positive relation between connectedness and the three cultural traits in analysis. However, due to the global dimension of our analysis, there can be a

significant heterogeneity of the effects across different regions in the world. We investigate the differential effects of connectedness on each cultural trait to understand whether the positive cultural effect of connectedness: (i) is bringing cultural convergence or divergence *across regions* (section 6.5.1), and (ii) is making populations *within regions* more or less culturally diverse (section 6.5.1).

Convergence and Divergence across regions

Given the positive relation of connectedness on each cultural trait, connectedness can generate cultural convergence or divergence across regions depending on the differential effect of connectedness compared to the average regional level of culture. On one hand, if the positive effect of connectedness is enhanced in regions which share a high level of cultural trait and is undermined in regions characterized by a low level of cultural trait, then being connected with people abroad will bring cultural divergence across regions. For instance, individuals living in regions where gender-egalitarian views are broadly shared may be more open to accept and adopt even more gender-equal attitudes shared by their connections abroad, compared to individuals who live in regions where gender-unequal views are common sense and are less prone to adopt uncommon views. In that case, connectedness makes regions culturally more distant. On the other hand, if the cultural effect is magnified in regions with a low level of cultural trait and minimized in regions with a high level of cultural trait, then connectedness will bring cultural convergence across regions. In that case, the effect may be driven by the information novelty introduced by the connection compared to the average norm. Using macro-data on migration, Rapoport *et al.* (2018) show that international migration is bringing overall cultural convergence across countries, although the direction of the convergence is still unclear. They suggest two different mechanisms: either through a convergence of host countries towards home countries, due to social mixing between migrants and natives (e.g. Baudin, 2010 and Chabé-Ferret, 2019), or a convergence of home countries towards host countries, driven by migrants social remittances at the origin (e.g. Levitt, 1998; Chauvet and Mercier, 2014; Docquier et al., 2016 and Valette, 2018).

To provide evidence of the latter, we split our sample of regions by terciles of average regional culture and perform a subsample analysis over each tercile of the distribution. Moreover, we provide additional evidence of the cultural convergence/divergence dynamic of connectedness by estimating equation (6.2) after replacing the average culture of connection's country of residence (i.e. $\overline{Cult_d}$) with the cultural distance between the latter and the average regional culture of each respondent. Such distance is computed as difference between the two averages, and a positive sign associated to the estimated parameters implies that the cultural effect

increases with the increase of the cultural distance. Since the individual cultural trait is included and influenced by the regional average culture, endogeneity arises and caution is needed when interpreting these latter results. Table 6.9 presents the results by cultural terciles using the standard specification presented in equation (6.2) in Panel A, while the estimates associated with cultural distances are presented in Panel B.[46] Columns (1) to (3) present the estimates on social behavior, columns (4) to (6) on religiosity and columns (7) to (9) on gender-egalitarian views.

Table 6.9: Cultural Convergence or Divergence - Regional subsamples by regional culture

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	OLS	OLS	OLS	LPM	LPM	LPM	OLS	OLS	OLS
	2009-2012	2009-2012	2009-2012	2009-2012	2009-2012	2009-2012	2009-2011	2009-2011	2009-2011
	Social Behavior			*Religiosity*			*Gender-Egalitarian*		
Tertile Trait	1^{st}	2^{nd}	3^{rd}	1^{st}	2^{nd}	3^{rd}	1^{st}	2^{nd}	3^{rd}
Panel A - Average									
Network	0.174***	0.202***	0.213***	0.030***	0.013***	0.003	0.085***	0.021	0.015
	(0.008)	(0.011)	(0.012)	(0.005)	(0.004)	(0.002)	(0.019)	(0.017)	(0.020)
Interaction	0.079***	0.105***	0.095***	-0.013	-0.021*	0.005	0.063	0.054	0.050
	(0.024)	(0.028)	(0.020)	(0.016)	(0.011)	(0.004)	(0.044)	(0.050)	(0.046)
Observations	166746	123398	121223	126457	127389	138055	33427	36409	20403
Regions	821	518	758	719	701	645	210	302	224
Adj. R-Square	0.06	0.05	0.09	0.14	0.06	0.02	0.18	0.08	0.05
Panel B - Distance									
Network	0.170***	0.200***	0.229***	0.026***	0.015***	0.005***	0.078***	0.022	0.021
	(0.009)	(0.011)	(0.012)	(0.004)	(0.004)	(0.002)	(0.022)	(0.017)	(0.019)
Distance	0.044**	0.097***	0.076***	0.067***	0.060***	0.007	0.046	0.059	0.072
	(0.018)	(0.027)	(0.021)	(0.022)	(0.017)	(0.006)	(0.032)	(0.052)	(0.059)
Observations	166746	123398	121223	126457	127389	138055	33427	36409	20403
Regions	821	518	758	719	701	645	210	302	224
Adj. R-Square	0.06	0.05	0.09	0.14	0.06	0.02	0.18	0.08	0.05
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region F.E.	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (col. (1)-(3)), religiosity dummy (col. (4)-(6)) and gender-egalitarian index (col. (7)-(9)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. The sample of regions is splitted by cultural traits terciles, respectively: first tercile (col. (1),(4),(7)), second tercile (col. (2),(5),(8)) and third tercile (col. (3),(6),(9)). Panel A shows the estimates using the specification presented in equation (6.2), while Panel B replace the average culture of connection's country of residence with a measure of cultural distance.

[46] Tables A17 and A18 in the Online Appendix shows the results after splitting our sample of regions by median and quartiles of average regional culture. The results are in line with the ones presented in Table 6.9.

There are two main findings we can take from Table 6.9. First, the pro-social behavior effect of connectedness is persistent and unrelated with the regional level of social behavior. Having a connection abroad is always associated with an higher degree of positive interactions with different actors within the society. The coefficients associated with connectedness and the interaction term have similar magnitude across different regional subsamples, although the effect is slightly stronger in regions in the upper tercile of regional social behavior. Those results are also confirmed by taking into account cultural distances rather then the average culture at the destination country (Panel B). Second, the positive effect of connectedness on religiosity and gender-egalitarian views is stronger in regions where the average cultural trait is lower. Panel A shows that individuals in regions belonging to the first tercile of the distribution of each trait are strongly influenced by connectedness: having a connection abroad is associated with an increase of 3% on the probability of being religious and 0.085 standard deviations of gender-egalitarian views.[47] The cultural effect of connectedness fades away and becomes statistically not significant with an increase of the regional average culture. Similar evidences are presented in Panel B, when cultural distances are included. Those results suggests a cultural convergence across regions due to connectedness concerning religiosity and gender-egalitarian views.[48] Although for gender-egalitarian views a catching up effect of gender-unequal regions towards more gender-equal regions is surely positive and good, for religiosity there is still an open debate.[49]

Diversity and Homogeneity within regions

To investigate whether the presence of connected individuals within each region is associated with a cultural homogeneous or heterogeneous population, we use the measurement framework proposed by Desmet *et al.* (2017) and Desmet and Wacziarg (2018). Those papers provide few measures to compute the degree of cultural heterogeneity of a given cultural value (which they define as "meme") and also how much identity cleavages (e.g. education, gender, etc.) can explain the heterogeneity in memes.

[47]Since the interaction terms are not statistically significant, we do not take them into account in evaluating the magnitude of the effect.

[48]Splitting the sample on broad geographical units gives similar results. Table A19 in the Online Appendix presents the estimates after splitting the sample in eight broad geographical unit: Europe, Former Soviet Union, Asia, Latina America, Middle East and North Africa, Sub-Saharan Africa, North America and Oceania. While pro-social behavior effect of connectedness is spread in all the continent, the positive effect on religiosity and gender-egalitarian attitudes is stronger in continents characterized by a lower level of each cultural trait (see Figures 6.2(IIb) and 6.2(IIIb) for a comparison).

[49]Some authors points out the negative economic effect of religiosity (Benabou *et al.*, 2015 and Chase, 2014), however religiosity is also associated to in increase of subjective well-being (Campante and Yanagizawa-Drott, 2015).

Let's consider a cultural trait $t = \{SB, RE, GE\}$, which can take values $n_t = m_t, ..., M_t$, where m_t and M_t are respectively the minimum and the maximum value of each trait. We define x^{n_t} the share of the total population which hold the value n_t of cultural trait t. Given our set of regions $r = 1, ..., R$, then we can compute the regional overall heterogeneity of cultural trait t as a simple measure of cultural fractionalization:

$$CF_r^t = 1 - \sum_{n_t=m_t}^{M_t} (x_r^{n_t})^2. \tag{6.4}$$

Namely, CF_r^t measure the probability to draw randomly two individuals from the population of region r which hold a different value of cultural trait t. To evaluate how much of the overall heterogeneity can be explained by the divide between connected and not connected individuals, we compute a measure of regional cultural fixation F_{ST}. F_{ST} measures the share of overall heterogeneity which can be explained by the heterogeneity between defined groups or cleavages of the society. Considering the individual characteristic $g = 1, ..., G$ which determines groups $z_r^g = 1, .., Z_r^g$ in the population of region r (e.g. g can be connectedness, and the groups are connected and not connected individuals), we can compute the overall cultural heterogeneity of cultural trait t in group z^g in region r as follow:

$$CF_{z_r^g}^t = 1 - \sum_{n_t=m_t}^{M_t} (x_{z_r^g}^{n_t})^2 \tag{6.5}$$

where $x_{z_r^g}^{n_t}$ is the share of individuals in region r belonging to group z^g who hold value n_t of cultural trait t. After defining the share of each group z_r^g in the population of region r as $s_{z_r^g}$, we can compute the within-group heterogeneity of cultural trait t as a weighted average over groups:

$$CF_{g_r}^t = \sum_{z_r^g=1}^{Z_r^g} s_{z_r^g} CF_{z_r^g}^t \tag{6.6}$$

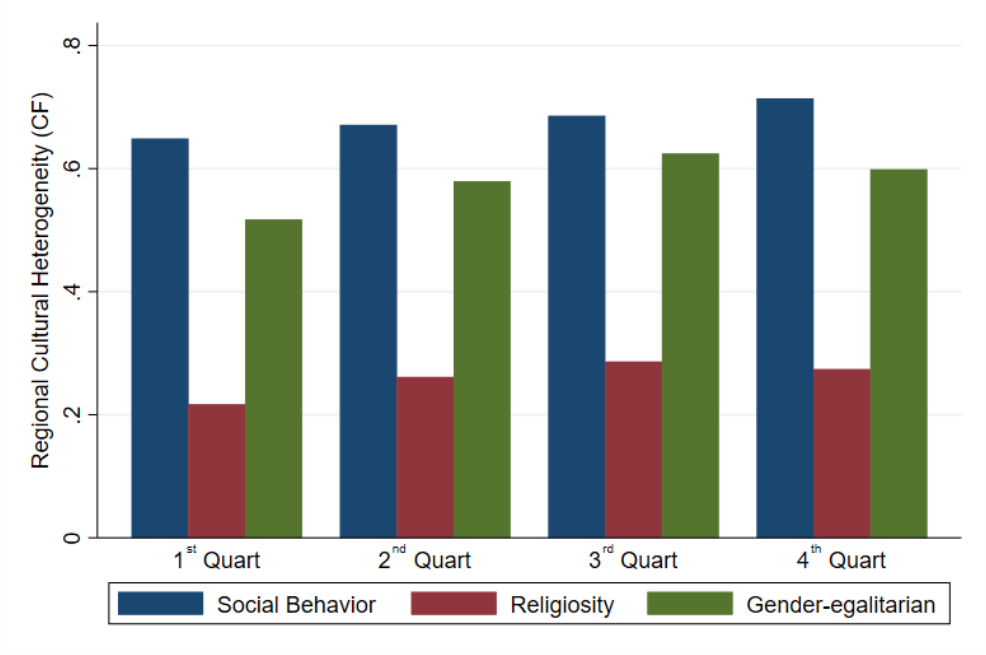
Finally, F_{ST} for a cultural trait t defined over the characteristic g in region r is just the share of the overall heterogeneity which is not explained by the within-group heterogeneity:

$$(F_{ST})_r^t = 1 - \frac{CF_{g_r}^t}{CF_r} \tag{6.7}$$

The index $(F_{ST})_r^t$ can takes values between 0 and 1. When it is equal to 0, the overall cultural

heterogeneity is explained by within-group heterogeneity. When it is equal to 1, the overall cultural heterogeneity is explained by the between-groups g heterogeneity (i.e., knowing the individual characteristic g is equal to know his/her cultural trait).

Figure 6.4: Within Regions Cultural Heterogeneity - Overall Heterogeneity

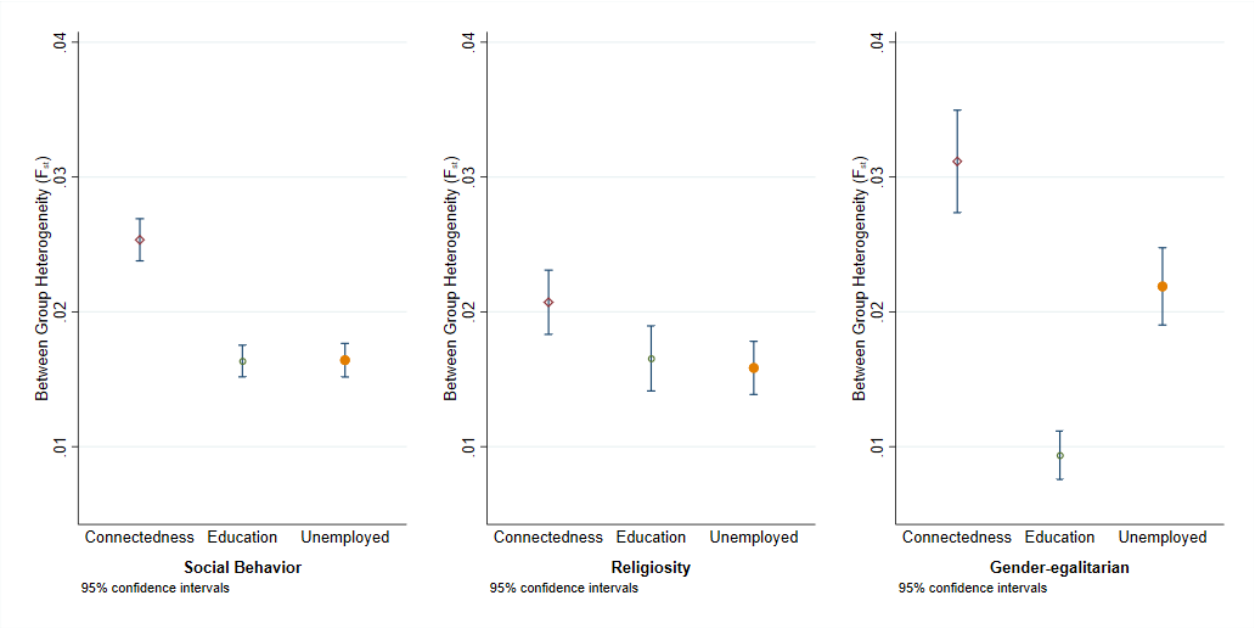


Note: authors' calculations on Gallup World Poll Data. The figure plots the average regional cultural heterogeneity computed over subsamples of regions based on share of connected people quartiles.

We first compute the overall regional heterogeneity for each cultural trait. We aggregate all the waves for each region, to have a time invariant measure of cultural heterogeneity for each trait. To investigate whether the presence of connected individuals influence the regional cultural diversity, Figure 6.4 shows the average regional cultural diversity of each trait computed over the subsamples of regions splitted by quartiles based on the share of connected people. Two are the main suggestive findings from those results. First, the overall regional diversity is trait specific. Social behavior and gender-egalitarian attitudes are characterized by an higher level of within-region diversity compared to religiosity. However, part of this effect is driven by nature of the variables: since religiosity is a dummy variable there is less variation in respondent answers, which will implies less variation in the shares used to compute the measure of cultural heterogeneity presented in equation (6.4). Second, the degree of within-region cultural diversity is positively related with the share of connected individuals in the region. The overall diversity is on average higher for each trait in region characterized by an higher presence of connected individuals. This result is a suggestive evidence that connectedness is contributing to an higher level of within-region cultural heterogeneity rather

then homogeneity.

Figure 6.5: Within Regions Cultural Heterogeneity - Between Groups Heterogeneity



Note: authors' calculations on Gallup World Poll Data. Each figure shows the average amount of total regional heterogeneity associated to each trait (CF^t) which is not attributable to within-group heterogeneity, but rather is related to between-group heterogeneity. Each region as the same weight. The defined groups-cleavages are based on: connectedness, education, employment status.

We then compute the amount of the overall regional cultural diversity which is not explained by the within-groups heterogeneity but rather then between-groups heterogeneity. Namely, we compute the F_{ST} index for each trait using connectedness as identity cleavage/group characteristic. Additionally, we compute the amount of overall heterogeneity explained by the between-groups heterogeneity based on other two characteristics: education and employment status. Desmet and Wacziarg (2018) show that those identity cleavages are the ones with the higher cultural divide, compared to other cleavages like gender or urbanicity. Figure 6.5 plots the average regional F_{ST} index over each cultural trait and characteristic. Overall, we find a low level of average regional cultural heterogeneity which is explained by between groups heterogeneity, which is on average between 2% and 3%. Those results suggest that cultural heterogeneity is not just explained by the cultural divide between some specific groups. Moreover, they are in line with Desmet and Wacziarg (2018), which find that the pluralism in the US is mainly explained by diversity within identity categories rather then diversity between identity categories. However, among the three individual characteristics in analysis, the cultural heterogeneity explained by between-groups diversity is higher when based on the connected/not connected divide. This is particularly true on social behavior

and gender-egalitarian attitudes, while less prominent for religiosity. Overall those results provide suggestive evidence that connectedness, compared to other relevant cleavages like education or employment status, identifies a specific set of cultural traits and values within regions.[50]

6.5.2 Mechanisms and channels of transmissions

In this section we exploit two sources of heterogeneity (regional and individual) to explore potential mechanisms and channel of transmissions of the cultural effect of connectedness. Section 6.5.2 tests whether the diffusion of connectedness within a region influence the magnitude of the individual cultural effect of connectedness. Section 6.5.2 explores whether individuals react differently to connectedness, depending on their education, gender, age and area of residence and whether connectedness affects individual economic outcomes. Finally, section 6.5.2 investigates whether there are differential effect of connectedness on connected individuals depending on the connection's country of residence.

Connectedness Diffusion

One might be concerned not only by cultural effects of having a connection abroad, but whether those effects are conditional to connectedness diffusion within a region. Regions characterized by an high diffusion of connectedness may be more used to receive and embrace information and novelties coming from their peers abroad, enhancing the cultural effect of connectedness due to a lower prejudice towards information coming from abroad. However, connected people living in regions where connectedness is not broadly diffused are exposed to unique information and novelties compare to other peers within the same region. In that case the effect of connectedness may be stronger in places where few people have a connection abroad. We empirically investigate those conditional cultural effects of connectedness by splitting our sample of regions by terciles of connectedness. We perform a subsample analysis on each tercile and we estimate our augmented equation for each cultural trait across different subsamples by connectedness diffusion.

Table 6.10 shows the estimates of our subsample analysis by connectedness diffusion. Columns (1) to (3) present the estimates on social behavior, columns (4) to (6) on religiosity and columns (7) to (9) on gender-egalitarian views. Overall the significance and the direction

[50]Table A20 in the Online Appendix provides a full set of statistics associated to the overall heterogeneity and between-group heterogeneity computed over the three cultural traits and the three individual characteristics.

Table 6.10: Connectedness Diffusion - Regional subsamples by connected population

	(1) OLS 2009-2012	(2) OLS 2009-2012	(3) OLS 2009-2012	(4) LPM 2009-2012	(5) LPM 2009-2012	(6) LPM 2009-2012	(7) OLS 2009-2011	(8) OLS 2009-2011	(9) OLS 2009-2011
	Social Behavior			*Religiosity*			*Gender-Egalitarian*		
Network Tertile	1^{st}	2^{nd}	3^{rd}	1^{st}	2^{nd}	3^{rd}	1^{st}	2^{nd}	3^{rd}
Network	0.218***	0.195***	0.181***	0.014***	0.012***	0.015***	0.003	0.061***	0.070***
	(0.012)	(0.010)	(0.010)	(0.005)	(0.004)	(0.004)	(0.022)	(0.018)	(0.019)
Interaction	0.160***	0.082***	0.100***	-0.010	-0.001	-0.009	0.150**	0.016	0.036
	(0.032)	(0.022)	(0.019)	(0.012)	(0.008)	(0.010)	(0.071)	(0.044)	(0.045)
Observations	156047	145555	109762	147009	137335	107555	37743	33043	19453
Regions	835	695	566	827	683	555	384	243	109
Adj. R-Square	0.17	0.14	0.16	0.44	0.33	0.36	0.22	0.19	0.26
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region F.E.	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (col. (1)-(3)), religiosity dummy (col. (4)-(6)) and gender-egalitarian index (col. (7)-(9)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. The sample of regions is splitted by terciles based on the regional share of connected individuals: first tercile (col. (1),(4),(7)), second tercile (col. (2),(5),(8)) and third tercile (col. (3),(6),(9)).

of the results are in line with our benchmark results. Moreover, connectedness diffusion does not significantly influence the cultural effect of connectedness on religiosity and social behavior. It is worth noticing that Beine et al. (2013) and Spilimbergo (2009) reached a similar conclusion in their papers on fertility norms and democracy, although looking at emigration rates rather than connectedness diffusion. The estimates remain reasonably close to each other across different subsamples, even though the effect are slightly stronger on individual social behavior when connectedness diffusion is low (column (1)). However, this is not the case for gender-egalitarian views. Comparing the estimates across different subsamples (columns (7)-(9)), the magnitude of the cultural effect of connectedness decreases with the share of the connected population. Column (7) shows that the effect becomes even not significant for individuals living in regions with the lowest diffusion of connectedness. Overall those results suggest that the diffusion of connectedness can influence the cultural effect of having a connection abroad, but this conditional effect is trait specific.[51]

[51]Tables A21 and A22 in the Appendix shows the results after splitting our sample of regions by median and quartiles of connectedness diffusion. The results are in line with the ones presented in Table 6.10.

Individual heterogeneity and economic channels

Individual characteristics of connected people could play a pivotal role to identify the channels through which connectedness is significantly related with culture. If one of the main effects of connections and networks is to affect the quality and size of individuals' information set (Granovetter, 2005 and Jackson, 2014), people with less information should react more to the new source of information. As the literature on political-persuasion has shown (Della Vigna and Gentzkow, 2010; Moriconi *et al.*, 2018), individuals revise their views based on their previous set of information in a Bayesian updating. The updating process is inversely related with the size of the information set: the more you know, the less you update. Pursuing this intuition, we split the population by education (more or less then 9 years of education), gender, location (rural and urban area) and age groups (15-35, 36-55 and 55+) and we estimate our equation (6.2) across different subsamples based on individual characteristics. Table 6.11 summarizes the results on social behavior (Panel A), on religiosity (Panel B) and on gender-egalitarian views (Panel C).

The effect of connectedness is positive and significant across the majority of the subsamples of the population. Moreover, the point estimates are quite close across different subsamples related to the same characteristic. However, although small, there are few differences across subsamples that are noteworthy. First, across education groups, the cultural effect of connectedness is not statistically significant or poorly significant on religiosity and gender-egalitarian attitudes among highly educated individuals. Those suggestive results point towards a potential mechanism of connectedness driven by the access to new information, which are more relevant for low educated individuals. However, given the imperfect measure of education available in GWP (i.e. whether individuals has at least nine years of education), the differential effect across education groups can be less precisely estimated. Second, the effect is stronger among males and among individuals in urban areas. Concerning the effects among the age groups, the results depends on the cultural trait in analysis. For instance, the gender-egalitarian effect of having a connection abroad increases with age. Compared to younger cohorts, individuals belonging to the age group 55+ are less used to gender-equal attitudes due to their recent evolution (Inglehart and Baker, 2000) and so the information brought by their connection abroad has an higher impact on their cultural values.

The cultural effect of connectedness can also go through some economic channels. If having a reliable connection abroad helps the economic condition of natives in the origin countries, then those natives can be more welcome to receive not only the economic aid but also the cultural beliefs of individuals abroad. First, we investigate whether having a reliable con-

Table 6.11: Individuals heterogeneity - Sociodemographic groups

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Education		*Gender*		*Age*			*Location*	
Individual Characteristic	LS	HS	Male	Female	15-35	36-55	55+	Rural	Urban
Panel A - Social Behavior (OLS)									
Network	0.195***	0.196***	0.209***	0.183***	0.191***	0.214***	0.183***	0.183***	0.200***
	(0.006)	(0.012)	(0.008)	(0.007)	(0.008)	(0.009)	(0.011)	(0.010)	(0.007)
Interaction	0.111***	0.035	0.094***	0.108***	0.094***	0.096***	0.108***	0.106***	0.095***
	(0.016)	(0.029)	(0.021)	(0.017)	(0.019)	(0.022)	(0.025)	(0.029)	(0.015)
Observations	357864	53356	187927	223437	193325	134647	84761	121278	290022
Regions	2096	1762	2094	2095	2085	2087	2016	1815	2038
Adj. R-Square	0.15	0.18	0.16	0.17	0.14	0.18	0.22	0.20	0.16
Panel B - Religiosity (LPM)									
Network	0.014***	0.012*	0.015***	0.013***	0.014***	0.018***	0.014***	0.009**	0.016***
	(0.002)	(0.007)	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)	(0.004)	(0.003)
Interaction	-0.007	-0.012	-0.006	-0.010	-0.010	-0.004	-0.005	-0.003	-0.009
	(0.006)	(0.017)	(0.008)	(0.007)	(0.007)	(0.009)	(0.012)	(0.008)	(0.007)
Observations	340524	51228	178510	213388	182049	128372	82833	116210	275624
Regions	2064	1731	2062	2064	2057	2057	1984	1784	2006
Adj. R-Square	0.38	0.31	0.40	0.36	0.40	0.41	0.32	0.39	0.37
Panel C - Gender-Egalitarian (OLS)									
Network	0.047***	0.042	0.052***	0.035**	0.041***	0.045**	0.091***	0.033*	0.051***
	(0.012)	(0.027)	(0.017)	(0.015)	(0.014)	(0.019)	(0.034)	(0.018)	(0.015)
Interaction	0.054*	0.022	0.091**	0.012	0.032	0.073	0.098	0.079	0.056
	(0.030)	(0.086)	(0.043)	(0.034)	(0.033)	(0.050)	(0.084)	(0.049)	(0.036)
Observations	83372	6779	44995	45242	54246	26419	9623	34151	56049
Regions	736	406	734	734	735	715	581	572	703
Adj. R-Square	0.22	0.24	0.21	0.20	0.22	0.24	0.25	0.24	0.23
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region F.E.	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (Panel A), religiosity dummy (Panel B) and gender-egalitarian index (Panel C). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. Each columns estimates the benchmark model in equation (6.2) by subsamples based on individual characteristics: education (col. (1)-(2)), gender (col. (3)-(4)), age group (col. (5)-(7)) and location (col. (8)-(9)).

nection abroad increase the probability to receive economic help from other individuals.[52] To minimize endogeneity driven by simultaneity bias, we include only controls which cannot be influenced by the dependent variable, namely gender and age. Column (1) of Table 6.12

[52]The GWP provide a question concerning the economic help provided to the household by other individuals. It asks as follow: "In the past 12 months, did this household receive help in the form of money or goods from another individual...?". Individuals can answer yes or no, and also they can specify whether they were helped by individuals living in the same country or abroad.

Table 6.12: Connectedness and Economic Outcomes

	(1) LPM	(2) LPM	(3) LPM	(4) LPM	(5) OLS	(6) OLS	(7) OLS	(8) OLS
Dep. Variable	Help	Help Local	Help Abroad	Unemployed	*Income*			
Sample	All	All	All	All	All	Help	Help Local	Help Abroad
Network	0.159***	0.002	0.126***	-0.002	1.616***	3.235	6.054	0.564***
	(0.005)	(0.002)	(0.005)	(0.001)	(0.531)	(2.332)	(4.818)	(0.104)
Female	0.013***	0.008***	0.004***	-0.006***	-0.277	2.604	4.873	-0.194***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.356)	(2.813)	(5.127)	(0.062)
Age	0.000	-0.000***	0.000***	-0.002***	-0.004	-0.066	-0.111	0.006***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.020)	(0.074)	(0.124)	(0.002)
Observations	486754	486754	486754	475792	481496	76731	46711	22925
Regions	2217	2217	2217	2201	2151	2034	1951	1332
Adj. R-Square	0.15	0.09	0.14	0.05	-0.00	-0.02	-0.04	0.16
Region F.E.	✓	✓	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is a dummy equal to one if the individual: received economic help (col. 1), received economic help from locals (col. 2), received economic help from individuals abroad (col. 3), is unemployed (col. 4). The dependant variable is the household income in thousands of international dollars from columns (5) to (8). The analysis is performed over the following samples: overall population (col. (1)-(5)), who receive economic help (col. 6), who receive economic help from locals (col. 7) and who receive economic help from abroad (col. 8).

shows that having a reliable connection abroad increase the probability to be economically helped by others by 16%. Moreover, when we split the dependent variable between individuals who receive economic help by individuals living in the same country (col. 2) or by individuals living abroad, as a proxy of remittances (col. 3), connectedness increase the probability to receive remittances around 12.5%, while it has no effect on the probability to receive help from other locals. Those results suggest a direct economic effect of connectedness on natives through remittances. Interestingly, connectedness does not affect other economic outcomes which are not directly related with having a reliable connection abroad, like the employment status (col. 4). Finally, we investigate the impact of connectedness on the household income from column (5) to (8) of Table 6.12. Having a reliable connection abroad is positively related with household income (col. 5). However, when we focus on the sample of individuals which are economically helped (col. 6), economically helped by locals (col. 7) or by people abroad (col. 8), the positive and statistically significant effect of connectedness on household income is estimated only on the latter group of individuals. This result is in line with the previous ones: connectedness influence the household level of income through remittances and not through other channels. Connectedness has not only a cultural effect, but also an economic effect on individuals in the origin countries, which can

act as channel of transmission of the cultural effect.

Differential effects based on connection location

The set of results presented in Section 6.4 shows that connected individuals are, on average, characterized by an higher level of social behavior, religiosity and gender-egalitarian attitudes, compared to not connected individuals living in the same region. However, due to the global scope of our analysis, connectedness can have differential effects on connected individuals depending also on connection's characteristic. In this section we focus on the sample of connected individuals, and we exploit the only reliable available information on connections by differentiating the connectedness variable on connection's country of residence. The country of residence can be a proxy of the set of culture and values which the connection has to face and interact with, and can proxy also the set of culture and values that the connection abroad communicate and share with natives in the origin countries.

First, since developed societies hold a distinctive set of culture and values (Inglehart, 2018), we explore the differential effect of having a reliable connection in an OECD high income countries or in another country. The estimates are presented in Panel A and B of Table 6.13. Distinctive effects are associated to connection's country of residence: having a connection in an OECD high-income country is associated with a strong positive effect on social behavior and gender-egalitarian attitudes, and a negative effect on religiosity. The estimates associated to a not OECD connection goes in the opposite direction but with a similar magnitude.

Another relevant differential effect can be driven by the majoritarian confession of connection's country of residence. The average confession does not influence only religiosity, but also a broad set of individual values and beliefs, from fertility norms to food consumption (Atkin *et al.*, 2019). For this reason, we evaluate the differential effect of having a reliable connection in a Muslim-majority (Panel C) or Christian-majority country (Panel D) on the sample of connected individuals.[53] Having a connection in a Muslim-majority country compared to having a connection in another country is associated with a lower effect on social behavior and gender egalitarian attitudes, while a stronger (but not precisely estimated) positive effect on religiosity. Having a connection in a Christian-majority country, instead, produce opposite results, although they are less precisely estimated.

[53]We define a Muslim-majority country if more than 50% of the population is considered as Muslim, following the Lugo and Cooperman (2011) data. A Christian-majority country is characterized by at least 50% of the population as Christian, using data from the CIA World Factbook, PEW Center and Joshua Project.

Table 6.13: Benchmark Results among Connected People - Connection Location

	(1) OLS 2009-2012	(2) OLS 2009-2012	(3) LPM 2009-2012	(4) LPM 2009-2012	(5) OLS 2009-2011	(6) OLS 2009-2011
	Social Behavior		*Religiosity*		*Gender-Egalitarian*	
Panel A - OECD Connection						
$Network^{OECD}$	0.075***	0.047**	-0.008	-0.030***	0.117***	0.123
	(0.017)	(0.019)	(0.006)	(0.011)	(0.041)	(0.109)
$Interaction^{OECD}$		0.067***		0.047**		-0.007
		(0.019)		(0.018)		(0.110)
Observations	120564	120564	115409	115409	24738	24738
Regions	2032	2032	1999	1999	643	643
Adj. R-Square	0.16	0.16	0.35	0.35	0.23	0.23
Panel B - Not OECD Connection						
$Network^{NOECD}$	-0.074***	-0.071***	0.009	-0.017	-0.121***	-0.122***
	(0.017)	(0.018)	(0.006)	(0.018)	(0.041)	(0.042)
$Interaction^{NOECD}$		0.025		0.036		0.039
		(0.050)		(0.023)		(0.098)
Observations	120564	120564	115409	115409	24738	24738
Regions	2032	2032	1999	1999	643	643
Adj. R-Square	0.16	0.16	0.35	0.35	0.23	0.23
Panel C - Muslim Maj. Connection						
$Network^{ISL}$	-0.069***	-0.069***	0.017*	0.108	-0.101*	-0.101*
	(0.024)	(0.024)	(0.010)	(0.066)	(0.058)	(0.057)
$Interaction^{ISL}$		0.056		-0.102		0.005
		(0.063)		(0.071)		(0.124)
Observations	120564	120564	115409	115409	24738	24738
Regions	2032	2032	1999	1999	643	643
Adj. R-Square	0.16	0.16	0.35	0.35	0.23	0.23
Panel D - Christian Maj. Connection						
$Network^{CHR}$	0.056***	0.024	0.003	-0.022**	0.107**	0.031
	(0.019)	(0.019)	(0.007)	(0.011)	(0.046)	(0.076)
$Interaction^{CHR}$		0.083***		0.051***		0.106
		(0.017)		(0.015)		(0.075)
Observations	120564	120564	115409	115409	24738	24738
Regions	2032	2032	1999	1999	643	643
Adj. R-Square	0.16	0.16	0.35	0.35	0.23	0.23
Region F.E.	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. Columns (1), (3) and (5) show the estimates from the specification presented in equation (6.1) while columns (2), (4) and (6) show the estimates from the specification presented in equation (6.2). The dependent variable is: social-behavior index (col. (1)-(2)), religiosity dummy (col. (3)-(4)) and gender-egalitarian index (col. (5)-(6)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income.

Overall, those estimates provide evidence that, after being overall related with an increase of all the cultural traits in analysis, connectedness has differential effects driven by the set

of culture and values of connection's country of residence.[54]

6.6 How much does it count?

Our results provide a new understanding of the individual cultural effect of having a connection in a foreign country, which lives and shares novel experience and information. Particularly, we show that connectedness makes individuals more active in the society, with an increase in their social behaviors, more religious and with a more open-minded attitude towards the role of women. Although the effect is positive and robust across different specifications and methodologies, the question about the relevance and the size of this effect still remain: is it a big or a small effect? In this section we produce some simulations of the effects based on our estimates. We first simulate the effect at the individual level and then we simulate the global effect.

At the individual level we consider, as a benchmark, the predicted culture of the average not connected individual by the coefficients estimated in Table 6.3. Using the same estimates, we predict the average culture of connected individuals, disregarding connection's location, and also accounting for the interaction effect. Lastly, for each cultural trait we compute the relative distance (in percentage) using not connected individuals as benchmark. Figure A8 in the Online Appendix shows the simulated individual relative cultural distance due to connectedness on social behavior, religiosity and gender-egalitarian views. Those simulations display two main results. First, average connected individuals have a distinctive set of cultural traits compared to not connected individuals. The size of relative distance varies significantly across cultural traits: social behaviors almost triples for connected individuals compared to not connected ones, while is around 37% for gender-egalitarian attitudes and around 10% for the probability of being religious. Second, the inclusion in the predictions of the effect driven by the interaction with the average culture of connection's country of residence has only an additional marginal effect.

After simulating the effect between not connected and connected individuals, we produce some simulations at the global level. As a benchmark for each cultural trait we consider the world level of culture when each country of the world is composed only by not connected

[54]However, it is important to recall that those effects are differential and on top of the overall effect of connectedness, while the main effect is driven by the difference between connected and not connected individuals. Tables A25 and A26 present estimates after splitting the sample of connections by the median or terciles of connection's country of residence culture. The results are not varying across subsamples, suggesting that the average cultural effect between connected and not connected individuals does not depend on the culture of connection country of residence.

individuals. We compute the benchmark as a weighted average of the culture of not connected people in each country, using as weight countries' population. We define such benchmark as follow:

$$\overline{Cult}_0^W = \frac{1}{Pop^W} \sum_{o \in W} \overline{Cult}_{NC}^o * Pop^o \tag{6.8}$$

where $\overline{Cult}_{NC}^o$ and Pop^o are the average culture of not connected people and the population in 2010 in country o, respectively. Then we compute the world cultural effect of connectedness by firstly defining, for each country, the share of connected people as γ^o. Assuming that connectedness is the only determinant of the cultural difference between connected and not connected individuals, and defining $\hat{\beta}_C$ the estimated coefficient of the effect of having a connection abroad on culture, we predict the global effect of connectedness as follow:

$$\overline{Cult}_1^W = \frac{1}{Pop^W} \sum_{o \in W} \left(\overline{Cult}_{NC}^o + \gamma^o \hat{\beta}_C\right) * Pop^o. \tag{6.9}$$

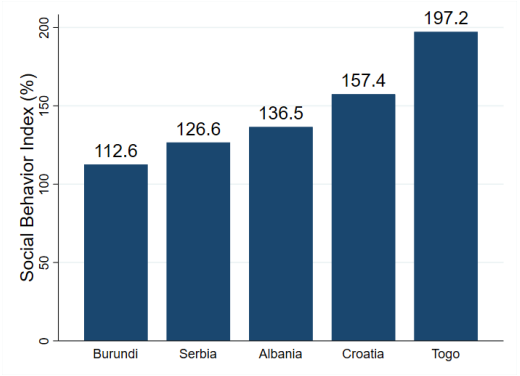
We also predict the level of culture after including the interaction effect with the culture of connection's country of residence. We compute, for each country o the average culture of connections' country of residence (defined as $\overline{Cult}_d^o$). Defining $\hat{\beta}_C^{Int}$ the estimated coefficient associated with the interaction term in equation (6.2), we predict the cultural effect of connectedness including the effect driven by connection's country of residence as follow:

$$\overline{Cult}_2^W = \frac{1}{Pop^W} \sum_{o \in W} \left[\overline{Cult}_{NC}^o + \gamma^o(\hat{\beta}_C + \hat{\beta}_C^{Int} * \overline{Cult}_d^o)\right] * Pop^o. \tag{6.10}$$

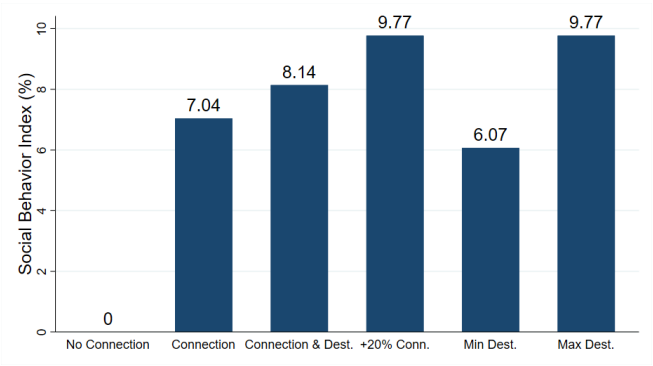
Figures 6.6(Ib), 6.6(IIb) and 6.6(IIIb) show the predicted global changes of each cultural trait, by computing the relative deviation between the benchmark (i.e. a world without connected people) and the predicted values computed in equations (6.9) and (6.10). Moreover they also display the world cultural change due to connectedness under three hypothetical scenarios: increasing the share of connected people by 20% within each country and replacing the average culture of connections' country of residence (namely $\overline{Cult}_d^o$) with the minimum and maximum computed level across all countries of the world.

Figure 6.6(Ib) shows the simulation on the social behavior index. The predicted global effect of connectedness on social behavior generates an increase around 7.04/8.14% compare to the benchmark scenario. Among the hypothetical scenarios, increasing the country share of connected individuals or associating to all the countries the highest value of social behavior of connections' country of residence give similar results, generating an increase of the global level of social behavior around 9.77% compare to the benchmark scenario. However, it is important to recall that there is a lot of cross-country heterogeneity, as it is reported in Table

Figure 6.6: Relative deviation from the No-connection Scenario - Country Level

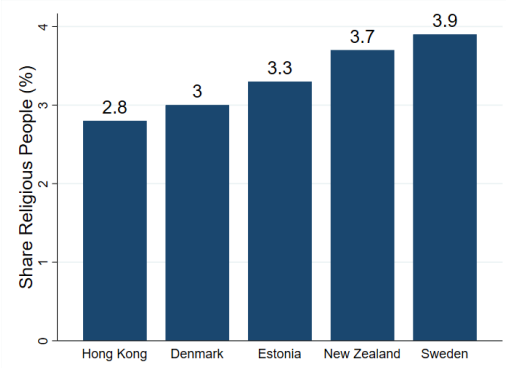


(a) Top-5 Countries

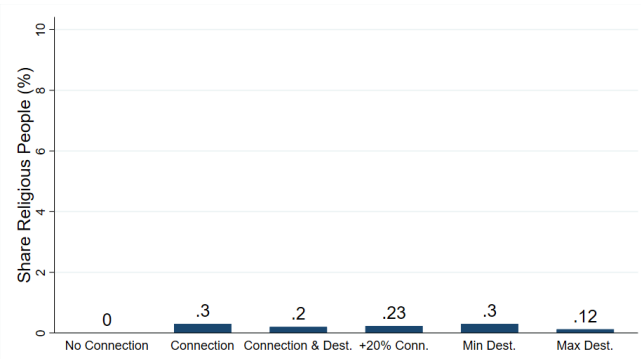


(b) World

(I) Social Behavior

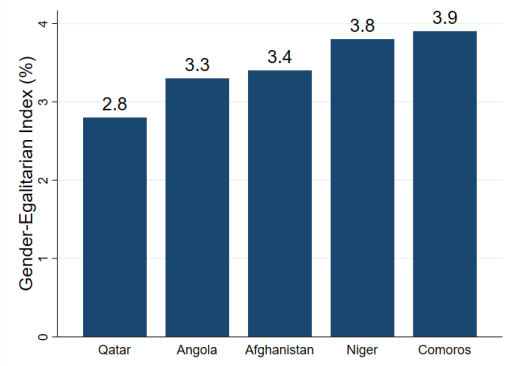


(a) Top-5 Countries

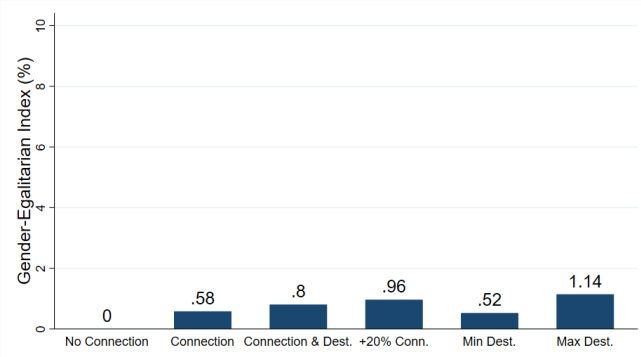


(b) World

(II) Religiosity



(a) Top-5 Countries



(b) World

(III) Gender-egalitarian views

Note: authors' calculations on Gallup World Poll Data. The Figure plots the simulated cultural effect of connectedness on each cultural trait. Figures (Ia), (IIa) and (IIIa) shows the top 5 countries that perform the highest relative distance compare to the no-connection scenario. Figures (Ib), (IIb) and (IIIb) show the simulated cultural effect of connectedness at the world level using the estimates presented in Table 6.3. Moreover it simulated the effects when the size of connected population increase of 20% in each country, and when all the connections reside in the country with the lowest and highest level of cultural trait.

A27 in the Online Appendix. Figure 6.6(Ia) shows the top 5 countries in terms of relative deviation from the benchmark scenario (i.e. a country without connected individuals): for those countries the distance is sizeable and it goes from 110% (Burundi) to 197% (Togo). Those sizeable effects are explained by both a low initial value of social behavior and an high share of connected individuals.

Concerning religiosity, the predicted global effect of connectedness is smaller. Figure 6.6(IIb) shows that it accounts for an increase of the share of religious people around 0.3% compared to the benchmark scenario, and it reaches it maximum level for few countries like Sweden and New Zealand, around 3.5/3.9% (see 6.6(IIa)). Increasing the size of the connected population or the average culture of connections' country of residence does not influence significantly the predictions, which remain small at the global level.

Figure 6.6(IIIb) displays the predicted global effect on gender-egalitarian attitudes. Gender-egalitarian attitudes relative distance compared to the benchmark scenario is around 0.58/0.8%. This predicted effect is smaller compared to the predicted effect on social behavior. Among the hypothetical scenarios, the scenarios associated with the extreme culture of connections' country of residence plays a significant role. Replacing $\overline{Cult}_d^o$ with the minimum and the maximum values in the sample, displayed by Tajikistan(-0.103) and Algeria(0.941), the predicted global increase in gender-egalitarian attitudes varies between 0.52% and 1.14%. In line with the simulations at the individual level, the pro-gender-egalitarian attitudes effect of having a connection abroad is significantly influenced by the culture of connection's country of residence. Among the countries mostly influenced by connectedness, Figure 6.6(IIIa) shows that Comoros and Niger are the countries which will benefit the most from their connection abroad.

Nevertheless, the results presented in Figure 6.6 should be interpreted cautiously and with few caveats in mind. Those results implicitly assume that all other individuals and regional factors are not changing or influencing the average culture of countries. In particular, if the increase of the share of connected individuals is due to emigration flows from countries of origin, then the average culture of the population can change if emigrants are culturally selected.[55] However, as Figure A9 in the Online Appendix shows, such specific issue should not overemphasized: the average culture of people in the origin countries and their diaspora abroad is similar across cultural traits.

[55]For instance, Anelli and Peri (2017) shows that Italian municipalities characterized by large outflows of young individuals experiences a deterioration of both turnout and vote for anti-establishment parties, due to a deterioration of political values hold by young cohorts.

6.7 Conclusions

In this paper we investigate the effect of having a reliable connection abroad on relevant individual cultural traits. Using an unique database on connectedness and individual beliefs, we test whether having a connection with individuals living abroad makes individuals culturally different compared to peers living in the same region. We first estimate at the individual level the aggregate effect across regions, and then we analyze heterogeneous effects, potential cultural convergences and divergences across and within regions and potential mechanisms and factors which can influence such cultural effect. To mitigate the endogeneity issue related to selection, we provide estimates after controlling for broader measure of connectedness, matching methods and testing for the potential threat driven by unobserved variables.

We address these questions using a large global sample of individuals over 148 countries, measuring both whether they have reliable connections abroad and some relevant cultural traits: social behavior, religiosity and gender-egalitarian views. We investigate the cultural effect of connectedness by comparing natives within the same intra-country region, and we run fixed-effects regressions for each distinctive cultural trait accounting for the culture of connections' country of residence. We strongly mitigate potential endogeneity threats due to selection into connectedness driven by observable characteristics by controlling for a broad measure of connectedness and through matching techniques. We also test the relevance of the threat driven by selection into unobservable characteristics. We test whether the effects vary across regions cultural level, bringing cultural convergence or divergence across and within regions. Moreover, we investigate potential mechanisms driven by the regional diffusion of connectedness, individual characteristics, economic channels and differential effects driven by connections location.

We find a robust and positive effect of having a connection abroad on natives social behavior, religiosity and gender-egalitarian attitudes. However, the size of the effect is different and trait specific. Those results are robust to the inclusion of broad measure of connectedness, matching techniques and selection on unobserved factors. The effect of connectedness is stronger in regions characterized by a low level of religiosity and gender-egalitarian views, suggesting a converging catch up effect across regions, while it increases cultural heterogeneity within regions. Concerning the mechanisms, the positive effect of connectedness is precisely estimated among less educated natives, which is consistent with the intuition that connections abroad are source of new information and experiences, which are more incisive for individuals who know less. Connectedness positively influences individuals economic conditions through economic remittances, which can also convey novel cultural norms.

Moreover, the location of the connection produce differential cultural effects on connected individuals, suggesting a transmission of destination-specific norms and traits.

We finally present some simulations to evaluate the magnitude of the cultural effect both at the individual and global level. At the individual level, connected individuals are significantly different compare to not connected ones. At the global level, the sizeable cultural effect of connectedness is perceivable on social behavior, while a minor, albeit not negligible, effect is predicted on gender-egalitarian and religiosity. Compared to a benchmark scenario where people have no connection abroad, the predicted cultural effect due to connectedness accounts at global level for a relative distance from the benchmark of 8.1% for social behavior, 0.8% for gender-egalitarian views and only 0.2% for religiosity.

This article shows that having reliable connection abroad makes individuals more active in the society, more religious and open-minded towards the role of women. However, further data on the experience and beliefs of the connection abroad should be needed to improve the understanding of the cultural effect of connectedness. We hope this study will stimulate further research on the determinants of individual culture and the interaction with their networks abroad.

6.8 Appendix

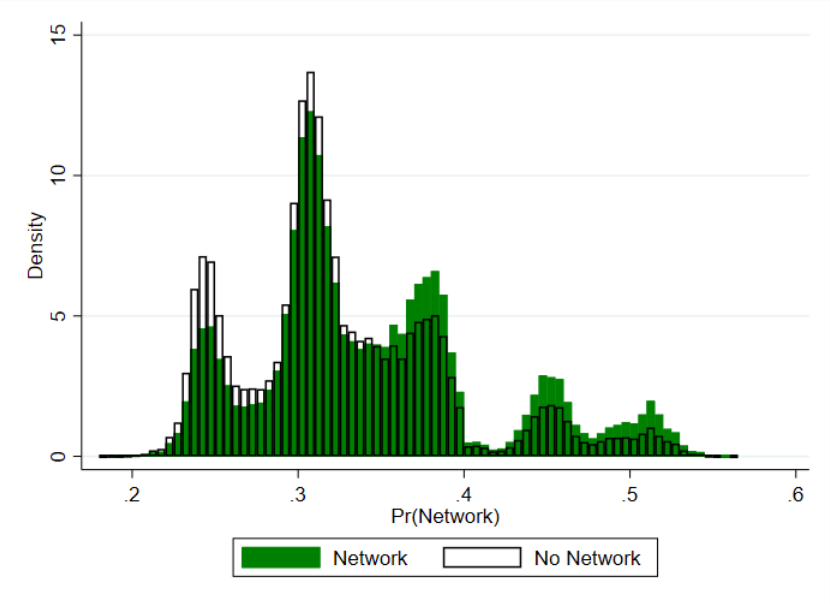
6.8.1 Matching: Tables and Figures

Table A1: Propensity Score Matching Estimation

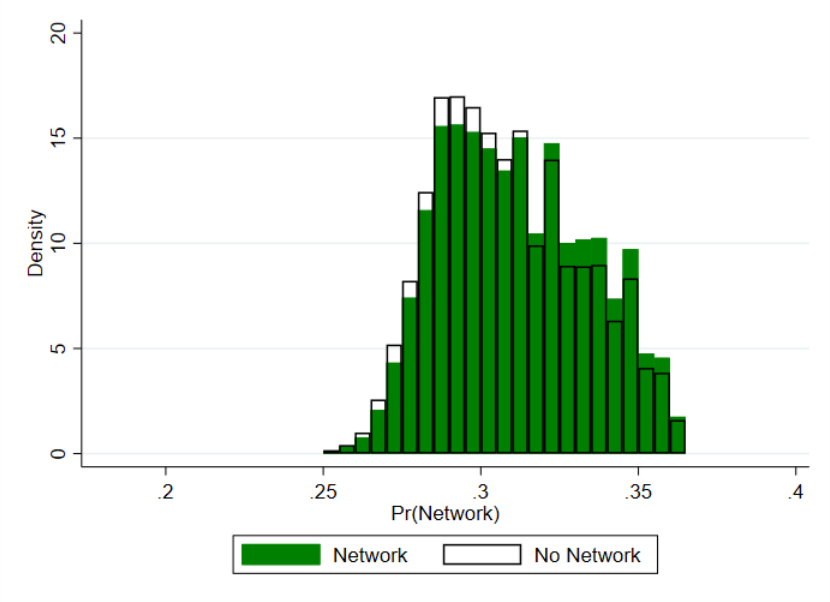
	(1) Probit 2009-2015	(2) Probit 2009-2015 Main	(3) Probit 2009-2015 Short
Education	0.3745***	0.3774***	
	(0.0050)	(0.0050)	
Female	-0.0262***	-0.0264***	-0.0248***
	(0.0034)	(0.0034)	(0.0030)
Married	-0.1249***	-0.1149***	
	(0.0036)	(0.0038)	
Age	-0.0025***	-0.0159***	-0.0144***
	(0.0001)	(0.0018)	(0.0015)
Urban	0.1963***	0.1936***	
	(0.0038)	(0.0038)	
Unempl.	0.0404***	0.0429***	
	(0.0068)	(0.0067)	
Family Size	0.0050***		
	(0.0009)		
Child	0.0058		
	(0.0037)		
Age^2		0.0003***	0.0002***
		(0.0000)	(0.0000)
Age^3		-0.0000***	-0.0000***
		(0.0000)	(0.0000)
Observations	839707	850609	1108028
Pseudo R2	0.01	0.01	0.00

Note: authors' calculations on Gallup World poll data. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is a dummy equal to one if the individual has a reliable connection abroad. The coefficient are estimated with a Probit model.

Figure A1: Probability Score Matching Distribution



(a) Main Model



(b) Short Model

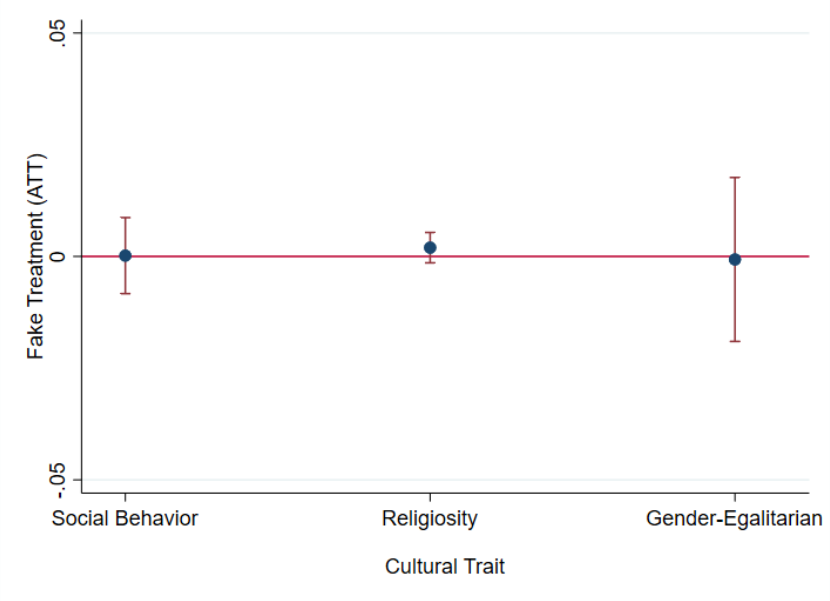
Note: authors' calculations on Gallup World Poll Data. The Figure plots the probability density of having a reliable connection abroad using the main model (figure a) and the short model (figure b) estimated in Table A1.

Table A2: Sample Means and Standardized Bias - Before and After Matching

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	All Sample			*Matched Sample (PS-Main)*			*Matched Sample (PS-Short)*			*Matched Sample (Mahala)*		
	Treated	Control	Bias (%)	Treated	Control	Bias (%)	Treated	Control	Bias (%)	Treated	Control	Bias (%)
Panel A - Social Behavior												
Education	0.14	0.08	11.65	0.18	0.17	0.75	0.18	0.13	9.42	0.18	0.17	1.14
Female	0.51	0.52	-1.47	0.55	0.56	-1.22	0.54	0.55	-0.92	0.55	0.55	-0.08
Married	0.49	0.55	-9.41	0.50	0.49	1.63	0.49	0.53	-4.98	0.49	0.50	-0.27
Child	0.58	0.59	-1.05	0.55	0.55	0.46	0.55	0.56	-1.62	0.55	0.55	-0.70
Age	37.07	38.91	-6.99	39.01	40.48	-5.58	39.00	40.21	-4.59	39.00	38.95	0.18
Urban	0.73	0.65	11.73	0.76	0.76	-0.79	0.75	0.73	3.33	0.76	0.76	-0.36
Unempl.	0.07	0.07	2.41	0.07	0.07	0.37	0.07	0.07	0.16	0.07	0.07	0.56
Family Size	3.69	3.65	1.28	3.29	3.21	2.75	3.29	3.21	2.89	3.29	3.26	0.98
Income	6.24	4.47	0.45	9.73	6.85	0.70	9.74	6.57	0.80	9.84	6.78	0.74
Panel B - Religiosity												
Education	0.14	0.08	11.52	0.18	0.18	0.80	0.18	0.13	9.62	0.18	0.17	1.18
Female	0.51	0.52	-1.46	0.55	0.56	-1.29	0.55	0.55	-0.95	0.55	0.55	-0.13
Married	0.48	0.55	-8.66	0.49	0.48	1.77	0.49	0.53	-4.91	0.49	0.49	-0.25
Child	0.58	0.59	-1.27	0.55	0.54	0.64	0.54	0.56	-1.45	0.54	0.55	-0.56
Age	37.27	39.04	-6.72	39.26	40.79	-5.80	39.25	40.50	-4.76	39.25	39.21	0.15
Urban	0.73	0.65	10.86	0.75	0.76	-0.86	0.75	0.73	3.36	0.75	0.76	-0.38
Unempl.	0.07	0.07	2.22	0.07	0.07	0.49	0.07	0.07	0.23	0.07	0.07	0.57
Family Size	3.66	3.64	0.78	3.26	3.17	2.89	3.26	3.17	2.94	03.26	3.22	1.15
Income	6.21	4.49	0.42	9.88	6.91	0.73	9.89	6.61	0.80	9.89	6.84	0.75
Panel C - Gender-Egalitarian												
Education	0.07	0.04	7.82	0.10	0.09	1.91	0.11	0.07	7.54	0.10	0.10	1.34
Female	0.50	0.51	-0.94	0.49	0.50	-1.14	0.50	0.51	-1.58	0.50	0.50	-0.01
Married	0.50	0.55	-6.12	0.50	0.50	0.35	0.49	0.53	-4.89	0.49	0.50	-0.35
Child	0.75	0.75	-0.13	0.73	0.74	-1.01	0.73	0.75	-2.85	0.73	0.74	-1.00
Age	33.43	34.41	-4.08	33.89	34.17	-1.13	33.89	33.84	0.21	34.89	33.48	1.70
Urban	0.65	0.58	8.21	0.67	0.68	-1.25	0.67	0.65	1.71	0.67	0.68	-0.67
Unempl.	0.08	0.07	0.95	0.08	0.08	-0.50	0.08	0.08	-0.56	0.08	0.08	1.03
Family Size	4.73	4.47	7.03	4.13	4.09	2.23	4.13	4.01	2.99	4.13	4.07	1.73
Income	2.56	2.10	2.34	3.18	2.82	1.80	3.19	2.53	3.29	3.20	3.22	-0.17

Note: authors' calculations on Gallup World poll data. * p<0.1, ** p<0.05, *** p<0.01. Columns (3), (6), (9) and (12) reproduce the standardized bias suggested by Rosenbaum and Rubin (1985). The standardized bias is computed as follow before and after the matching procedure: $SB_{Bef}(X) = 100 * \frac{\overline{X_1} - \overline{X_0}}{\sqrt{[V_1(X)+V_0(X)]0.5}}$ and $SB_{Aft}(X) = 100 * \frac{\overline{X_{1M}} - \overline{X_{0M}}}{\sqrt{[V_1(X)+V_0(X)]0.5}}$

Figure A2: Matching Results - Robustness to fake treatment on control group



Note: authors' calculations on Gallup World Poll Data. The Figure plots the average effect of a fake treatment drawn from an uniform distribution and randomly assigned over the not connected individuals belonging to the control group on three different cultural traits (Social Behavior, Religiosity and Gender-Egalitarian) and the interval of confidence at 99% level. The propensity score matching is done using Kernel Epanechnikov matching. Standard errors are bootstrapped.

6.8.2 Matching: Sensitivity Analysis

Following the sensitivity analysis proposed by Rosenbaum (2002), we test the robustness of our estimated cultural effects to unobserved factors after matching methods. Matching methods manage to eliminate bias driven by selection into observable characteristics, by minimizing the difference between connected and not connected individuals. However, as for linear methods, they are not robust against "hidden bias" (DiPrete and Gangl, 2004): unobserved factors that affect simultaneously individual culture and conncetedness. Even though section 6.4.2 shows that selection on unobservables is a minor concern, we decide to follow the matching literature (Aakvik, 2001 and DiPrete and Gangl, 2004) and compute the Rosenbaum bounds of our estimates.[56] Assumed a certain level of hidden bias (presented with the variable γ), such approach allows us to compute the bounds of average cultural effect of connectedness, once we assume that hidden bias is causing an over-estimation of the effect (MH^+) and under-estimation of the effect (MH^-). To give an intuition behind the value of γ, a value of $\gamma = 1$ is associated with no hidden-bias, while $\gamma = 1.5$ implies that individuals with the same characteristics differ in their odds of having a connection by a factor of 1.5. Following Becker and Caliendo (2007), we use Mantel and Haenszel (1959) test statistic to compute the Rosenbaum bounds after PSM. Table A3 presents the results of our sensitivity analysis on social behavior (panel A), religiosity (panel B) and gender-egalitarian views (panel C). Column (1) reports the level of hidden bias (γ). Since the cultural effect of connectedness is always positive, then our concern is related mainly to an over-estimation of the effect due to hidden bias. For this reason we report the p-value related to the over-estimated (MH^+) bound in column (2). Columns (3) and (4) report the bounds while column (5) the confidence interval. The critical level of γ at which our estimated positive cultural effect of having a connection abroad after matching should be questioned is between 1.85 and 1.90 for social behavior, between 1.20 and 1.25 for gender-egalitarian views and between 1.15 and 1.20 for religiosity. It is important to recall that such approach allow us to evaluate the robustness of our results in the *worst-case scenario*. For instance, concerning the effect on religiosity, the cultural effect of connectedness would include zero if unobserved variables influence the odds ratio of having a connection abroad between connected and not connected people with equal characteristics by 1.15 and influence the cultural trait. If unobserved factor influence only individual connectedness and not individual culture, then the confidence interval should not include zero. Nonetheless, those results suggest that the cultural effect of connectedness is less robust on religiosity, compared to gender-egalitarian

[56] Due to the high amount of regions in our analysis, the available Stata packages have issues to compute Rosenbaum Bounds over an high number of strata. For this reason, for this sensitivity analysis we perform the matching disregarding the geographical location of individuals.

views and social behavior.

Table A3: Sensitivity Analysis - Rosenbaum Bounds

	(1) Gamma (γ)	(2) p-value	(3) MH^+	(4) MH^-	(5) C.I.
Panel A - Social Behavior					
	1.00	<0.001	0.159	0.159	[0.158, 0.160]
	1.80	<0.001	0.010	0.608	[0.005, 0.609]
	1.85	<0.001	0.003	0.613	[0.002, 0.615]
	1.90	0.341	0.001	0.616	[-0.009, 0.617]
	1.95	0.999	-0.021	0.617	[-0.052, 0.618]
Panel B - Religiosity					
	1.00	<0.001	0.227	0.227	[0.227, 0.228]
	1.10	<0.001	0.224	0.234	[0.223, 0.235]
	1.15	<0.001	0.219	0.237	[0.217, 0.238]
	1.20	0.102	0.206	0.237	[-0.048, 0.239]
	1.25	1.00	-0.075	0.240	[-0.085, 0.241]
Panel C - Gender-Egalitarian					
	1.00	<0.001	0.217	0.217	[0.211, 0.231]
	1.10	<0.001	0.183	0.272	[0.161, 0.292]
	1.15	<0.001	0.156	0.304	[0.137, 0.311]
	1.20	0.013	0.117	0.313	[0.006, 0.318]
	1.25	0.780	-0.003	0.321	[-0.011, 0.332]

Note: authors' calculations on Gallup World poll data. Column (1) presents the difference in odds of having a connection abroad between matched connected and not connected individuals (hidden bias). Column (2) present the significance level associated to the over-estimated bound (MH^+). Columns (3) and (4) show the over-estimated and under-estimated bounds due to hidden bias. Column (5) shows the confidence interval of the estimated bounds.

6.8.3 Descriptive Statistics

Table A4: List of origin countries (157)

Afghanistan, Albania, Algeria, Angola, Argentina, Armenia, Australia, Austria, Azerbaijan, Bahrain, Bangladesh, Belarus, Belgium, Benin, Bolivia, Bosnia and Herzegovina, Botswana, Brazil, Bulgaria, Burkina Faso, Burundi, Cambodia, Cameroon, Canada, Central African Republic, Chad, Chile, China, Colombia, Comoros, Costa Rica, Croatia, Cyprus, Czech Republic, Côte d'Ivoire, Democratic Republic of the Congo, Denmark, Djibouti, Dominican Republic, Ecuador, Egypt, El Salvador, Estonia, Ethiopia, Finland, France, Gabon, Georgia, Germany, Ghana, Greece, Guatemala, Guinea, Haiti, Honduras, Hong Kong, Hungary, Iceland, India, Indonesia, Iran, Iraq, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kazakhstan, Kenya, Kosovo, Kuwait, Kyrgyzstan, Laos, Latvia, Lebanon, Lesotho, Liberia, Libya, Lithuania, Luxembourg, Macedonia, Madagascar, Malawi, Malaysia, Mali, Malta, Mauritania, Mauritius, Mexico, Moldova, Mongolia, Montenegro, Morocco, Mozambique, Myanmar, Nepal, Netherlands, New Zealand, Nicaragua, Niger, Nigeria, Northern Cyprus, Norway, Oman, Pakistan, Palestina, Panama, Paraguay, Peru, Philippines, Poland, Portugal, Qatar, Republic of Congo, Romania, Russia, Rwanda, Saudi Arabia, Senegal, Serbia, Sierra Leone, Singapore, Slovakia, Slovenia, Somalia, South Africa, South Korea, Spain, Sri Lanka, Sudan, Suriname, Swaziland, Sweden, Switzerland, Syria, Taiwan, Tajikistan, Tanzania, Thailand, Togo, Trinidad and Tobago, Tunisia, Turkey, Turkmenistan, Uganda, Ukraine, United Arab Emirates, United Kingdom, United States, Uruguay, Uzbekistan, Venezuela, Vietnam, Yemen, Zambia, Zimbabwe

Table A5: Descriptive Statistics

Variable	(1) Obs	(2) Mean	(3) Std.Dev.	(4) Min	(5) Max
Network	573,425	0.314	0.464	0	1
Education	703,827	0.154	0.360	0	1
Female	712,722	0.535	0.499	0	1
Married	708,664	0.547	0.498	0	1
Child	701,653	0.537	0.499	0	1
Age	709,064	40.580	17.289	15	90
Urban	683,310	0.741	0.4378	0	1
Family size	698,058	3.217	1.880	0	78
Unempl.	657,520	0.060	0.238	0	1
Income	689,822	0.017	6.068	0	5014.626
Family Net.	435,037	0.033	0.178	0	1
Migration Int.	640,004	0.198	0.399	0	1
Migration Plan.	640,004	0.017	0.131	0	1
Migrant	712,724	0.053	0.223	0	1
Internet	692,959	0.369	0.482	0	1
Landline Phone	673,762	0.478	0.499	0	1
Social Behavior	626,329	0	1	-1.105	2.622
Religiosity	597,269	0.731	0.443	0	1
Gender-Egalitarian	141,077	0	1	-3.025	0.816

Note: authors' calculations on Gallup World Polls.

Table A6: Pearson's Correlation across cultural questions

Variable	(1) *Social Behavior*	(2) *Religiosity*	(3) *Gender-Egalitarian*
Social Behavior (ind.)	1.0000		
Religiosity (RE_1)	0.0420***	1.0000	
Gender-egalitarian (ind.)	-0.0046*	-0.0206***	1.0000
Donate money (SB_1)	0.7087***	-0.0045***	-0.0169***
Volunteer (SB_2)	0.7111***	0.0258***	0.0163***
Helped stranger (SB_3)	0.6781***	0.0685***	-0.0099***
Women same right as men (GE_1)	-0.0045*	-0.0273***	0.7583***
Women right to hold job (GE_2)	0.0052*	0.0043	0.7684***
Women right initiate divorce (GE_3)	-0.0082***	-0.0277***	0.6634***

Note: authors' calculations on Gallup World Polls. The table shows Pearson's correlations across indicators and questions related to culture. * p<0.1, ** p<0.05, *** p<0.01.

6.8.4 Active Religious participation and Discrimination

This section investigates even further the implication and potential causes of the cultural effect of having a connection abroad. The results presented in section 6.4 show that having a connection abroad is associated with higher social behavior, religiosity and gender-egalitarian views. However few questions still remain. The positive effect of connectedness on religiosity requires further explanations, to understand the implications for the society: does the pro-religiosity effect of connectedness impact individual direct implication to religious activities? Does it vary across religious groups? Moreover individuals can react differently to the information given by their connection due to their social context: does perceived discrimination towards social outcasts and women influence the cultural effect of connectedness on social behavior and gender-egalitarian views?

Table A7: Religiosity: Practicing and Religious groups

	(1) LPM	(2) LPM	(3) LPM	(4) LPM	(5) LPM	(6) LPM	(7) LPM	(8) LPM
	Practicing		Religiosity & Religious Groups					
	Place of Worship	Trust Rel. Org.	Christians	Muslims	Hinduists	Buddhists	Agnostics	Other Rel
Network	0.012	0.003	0.007**	0.020***	0.019	0.006	0.004	0.021
	(0.007)	(0.003)	(0.003)	(0.003)	(0.016)	(0.008)	(0.006)	(0.023)
Interaction	0.021	0.002	-0.020***	-0.006	-0.040	0.010	0.005	-0.041
	(0.013)	(0.006)	(0.007)	(0.009)	(0.040)	(0.020)	(0.020)	(0.042)
Observations	159756	325925	111046	191946	21387	18061	18556	6138
Regions	1657	1900	816	1625	80	239	816	436
Adj. R-Square	0.25	0.17	0.26	0.29	0.14	0.32	0.21	0.25
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓
Region F.E.	✓	✓	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: being to a place of worship dummy (col. (1)), trust religious organizations dummy (col. (2)) and religiosity dummy (col. (3)-(8)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. The sample of individual is splitted by religious groups between columns (3) and (8): Christians (col. (3)), Muslims (col. (4)), Hinduists (col. (5)), Buddhists (col. (6)), agnostics (col. (7)) and a residual group of other religions (col. (8)).

We address the issues related to the pro-religiosity effect of connectedness in two ways. First, we exploit two questions available in the GWP concerning active participation to religious life and confidence in religious organizations. The questions are the following: (i) "Have you attended a place of worship or religious service within the past seven days?" and (ii) "In this country, do you have confidence in religious organizations (churches, mosques, temples, etc.)?". They are coded as dummy variables which take value of one if individual answer af-

firmatively. These item is available only over 2009-2011 period and on a subset of countries. We use them as alternative dependent variable, to investigate the effect of connectedness on actual religious practices. Second, we perform a subsample analysis based on individual religious groups to understand whether the pro-religiosity effect of connectedness is enhancing individuals belonging to some specific religious groups or is affecting all individuals indiscriminately. Table A7 presents the estimates using the benchmark equation (6.2) as a model but with different dependent variables and on different subsamples. Columns (1) and (2) show the estimates on active religious participation and confidence towards religious organization, respectively, while columns (3) to (8) shows the cultural effect of connectedness on religiosity across different religious groups. Two main findings emerge from this table. First, the pro-religiosity effect of connectedness is not associated with an active implication with religious organizations: having connection abroad has no statistically significant effect on individual attendance at places of worship and on individual confidence towards religious organizations. Those results are in line with the arguments of Wuthnow (1998) and Inglehart *et al.* (2017), which point out that individual allegiance towards religious institutions is declining but not spiritual concerns. Second, the pro-religiosity effect is stronger among Christians and Muslims, and not statistically significant among other religious groups or among agnostic individuals. Such result suggest that the pro-religiosity effect of connectedness is reinforcing religiosity of individuals who already belong to a religious group, rather then affecting individuals who are far from a religious life.

Concerning the positive effect of connectedness on social behavior and gender-egalitarian views, individual's perception of the society may influence the size of the effect. Individuals who perceive an unequal society towards social outcasts and women could have a strong cultural reaction after interacting with the new source of information given by his/her connection abroad. We test whether perceived justice towards people/migrants and discrimination towards women influence the cultural effect of connectedness by performing a subsample analysis at individual level. Using three questions[57] available in GWP, we split the sample of individuals between who is satisfied on how poor people/migrants/women are treated in the country and who, instead, think that they are treated poorly or without respect. Table A8 presents the estimates using social behavior index (col. (1) to (4)) and gender-egalitarian index (col. (5)-(6)) as dependent variable. Overall, the estimates are positive and significant across all the subsamples, suggesting a stable and robust cultural effect of connectedness.

[57]The question associated to poor people is: "In this country, are you satisfied or dissatisfied with efforts to deal with the poor?". Concerning immigrants, GWP asks the following question: "Is the city or area where you live a good place or not a good place to live for immigrants from other countries?". Finally, concerning discrimination towards women: "Do you believe that women in this country are treated with respect and dignity, or not?".

Table A8: Social Behavior and Gender-egalitarian views: perceived discrimination

	(1) OLS 2009-2012	(2) OLS 2009-2012	(3) OLS 2009-2012	(4) OLS 2009-2012	(5) OLS 2009-2011	(6) OLS 2009-2011
Dep. Var.	*Social Behavior*				*Gender-Egalitarian*	
	Poor People		Migrants		Women	
Treated	Not Satisfied	Satisfied	Not Good Place	Good Place	Not Respect	Respect
Network	0.194***	0.214***	0.173***	0.199***	0.047**	0.051***
	(0.008)	(0.012)	(0.009)	(0.008)	(0.021)	(0.014)
Interaction	0.090***	0.082***	0.128***	0.087***	0.080*	0.017
	(0.018)	(0.024)	(0.028)	(0.018)	(0.046)	(0.042)
Observations	158702	92896	127527	207135	22715	59662
Regions	1919	1881	2043	2069	637	659
Adj. R-Square	0.17	0.16	0.17	0.17	0.27	0.22
Individual Controls	✓	✓	✓	✓	✓	✓
Region F.E.	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (col. (1)-(4)) and gender-egalitarian index (col. (5)-(6)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. Each regression includes regional and country-year fixed effects. The estimates are computed on the subsample of individuals who: is not satisfied/satisfied (col. (1)/(2))) with the effort to deal with the poor in the country, think that the city/area is not a good place/good place (col.(3)/(4)) to live for immigrants from other countries, believe that women are not treated/treated (col. (5)/(6)) with respect and dignity in the country.

Nevertheless, columns (5) and (6) show that the magnitude of the effect on gender-egalitarian views is influenced by perceived discrimination. Taking the estimates at their face value and using the average gender-egalitarian views in the destination countries (0.587), having a connection abroad increase gender-egalitarian views by 0.108(=0.042+0.113*0.587) standard deviation among individuals who think that women are not treated with respect in the country, while only 0.061(=0.056+0.010*0.587) standard deviation among individuals who do not perceive a discriminatory society towards women. Those results suggest that gender-egalitarian norms introduced by the interaction with peers abroad are stronger when discrimination towards women is perceived.

6.8.5 Additional Robustness

Table A9: Connectedness or Openness - Individual controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	OLS	OLS	OLS	LPM	LPM	LPM	OLS	OLS	OLS
	2009-2012	2009-2012	2009-2012	2009-2012	2009-2012	2009-2012	2009-2011	2009-2011	2009-2011
	Social Behavior			*Religiosity*			*Gender-Egalitarian*		
Additional Controls	Migration	Communication	Both	Migration	Communication	Both	Migration	Communication	Both
Network	0.192***	0.179***	0.175***	0.016***	0.017***	0.019***	0.045***	0.037***	0.037***
	(0.006)	(0.006)	(0.006)	(0.002)	(0.002)	(0.002)	(0.012)	(0.012)	(0.012)
Interaction	0.101***	0.089***	0.090***	-0.007	-0.009	-0.009	0.053*	0.054*	0.052*
	(0.013)	(0.013)	(0.013)	(0.006)	(0.006)	(0.006)	(0.030)	(0.031)	(0.031)
Observations	406441	402748	398806	387959	384266	380324	88305	87353	86403
Regions	2079	2083	2073	2055	2059	2049	723	722	717
Adj. R-Square	0.16	0.17	0.17	0.38	0.38	0.38	0.22	0.21	0.21
Migration Controls	✓		✓	✓		✓	✓		✓
Means of Communication		✓	✓		✓	✓		✓	✓
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region F.E.	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (col. (1)-(3)), religiosity dummy (col. (4)-(6)) and gender-egalitarian index (col. (7)-(9)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. As additional migration controls we include: migration status, willingness to migrate permanently abroad and planning to migrate permanently abroad. As additional means of communications controls we include: having internet and having land-line telephone in the house.

Table A10: Connectedness or Openness - Country subsamples by development and trade exposure

	(1) OLS 2009-2012	(2) OLS 2009-2012	(3) LPM 2009-2012	(4) LPM 2009-2012	(5) OLS 2009-2011	(6) OLS 2009-2011
	Social Behavior		*Religiosity*		*Gender-Egalitarian*	
Median Characteristic	Below	Above	Below	Above	Below	Above
Panel A - GDP per Capita						
Network	0.176***	0.222***	0.013***	0.018***	0.042***	0.061**
	(0.008)	(0.009)	(0.003)	(0.004)	(0.013)	(0.024)
Interaction	0.136***	0.058***	-0.009	-0.007	0.057	0.042
	(0.024)	(0.016)	(0.006)	(0.012)	(0.035)	(0.075)
Observations	225409	176333	212143	172054	69916	16445
Regions	960	1118	936	1110	526	191
Adj. R-Square	0.15	0.17	0.36	0.30	0.21	0.12
Panel B - Import						
Network	0.203***	0.193***	0.007*	0.020***	0.040**	0.054***
	(0.009)	(0.009)	(0.004)	(0.003)	(0.017)	(0.016)
Interaction	0.099***	0.096***	-0.004	-0.013	0.044	0.071
	(0.020)	(0.019)	(0.008)	(0.008)	(0.037)	(0.053)
Observations	209837	191241	203728	180792	51071	35225
Regions	1091	967	1089	939	497	209
Adj. R-Square	0.17	0.15	0.40	0.35	0.19	0.26
Panel C - Export						
Network	0.196***	0.200***	0.009***	0.019***	0.034**	0.065***
	(0.009)	(0.008)	(0.003)	(0.004)	(0.015)	(0.019)
Interaction	0.115***	0.080***	-0.004	-0.014	0.057*	0.058
	(0.020)	(0.019)	(0.007)	(0.010)	(0.034)	(0.068)
Observations	216810	184268	213189	171331	55663	30633
Regions	1105	953	1105	923	504	202
Adj. R-Square	0.17	0.15	0.41	0.33	0.24	0.19
Individual Controls	✓	✓	✓	✓	✓	✓
Region F.E.	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (col. (1)-(2)), religiosity dummy (col. (3)-(4)) and gender-egalitarian index (col. (5)-(6)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. The sample of countries is splitted by the median level of GDP per capita (Panel A), import (Panel B) and export (Panel C) as share of GDP: below the median (col. (1),(3),(5)) and above the median (col. (2),(4),(6)).

Table A11: Benchmark Estimates - Probit and Ordered Probt Model

	(1) Oprobit 2009-2012	(2) Probit 2009-2012	(3) Oprobit 2009-2011
	Social Behavior	*Religiosity*	*Gender-Egalitarian*
Network	0.234***	0.065***	0.057***
	(0.007)	(0.011)	(0.015)
Interaction	0.094***	-0.027	0.066*
	(0.015)	(0.029)	(0.035)
Observations	411367	385046	90239
Regions	2097	1938	736
Pseudo-R^2	0.05	0.34	0.08
Individual Controls	✓	✓	✓
Region F.E.	✓	✓	✓
Country-year F.E.	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (col. (1)), religiosity dummy (col. (2)) and gender-egalitarian index (col. (3)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. Coefficients are estimated using a Probit model (col. (2)) and Ordered Probit Model (col. (1) and (3)).

Table A12: Analysis on Subcomponents of Social Behavior and Gender-Egalitarian attitudes

	(1) LPM 2009-2012 Donate (SB_1)	(2) LPM 2009-2012 Volunteer (SB_2)	(3) LPM 2009-2012 Stranger (SB_3)	(4) LPM 2009-2011 Women as man (GE_1)	(5) LPM 2009-2011 Women Job (GE_2)	(6) LPM 2009-2011 Women Divorce (GE_3)
Network	0.058***	0.041***	0.082***	0.012**	0.014***	0.031***
	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)	(0.007)
Interaction	0.030***	0.065***	0.020**	-0.006	0.006	-0.005
	(0.008)	(0.013)	(0.009)	(0.014)	(0.010)	(0.016)
Observations	411367	410373	411367	90239	90239	85874
Regions	2097	2095	2097	736	736	712
Adj. R-Square	0.20	0.11	0.10	0.18	0.14	0.19
Individual Controls	✓	✓	✓	✓	✓	✓
Region F.E.	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variables are dummy whether an individual: donated money in the last month (col. 1), volunteered time to an organization (col.2) helped a stranger (col. 3), believe that women and men should have the same rights (col. 4), believe that women should have the right to hold a job (col. 5) and believe that women should have the right to initiate a divorce. The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income.

Table A13: Benchmark Analysis - Countries Land

	(1) OLS 2009-2012	(2) OLS 2009-2012	(3) OLS 2009-2012	(4) LPM 2009-2012	(5) LPM 2009-2012	(6) LPM 2009-2012	(7) OLS 2009-2011	(8) OLS 2009-2011	(9) OLS 2009-2011
	Social Behavior			*Religiosity*			*Gender-Egalitarian*		
Land KM^2	>50000	>100000	>200000	>50000	>100000	>200000	>50000	>100000	>200000
Network	0.205***	0.207***	0.216***	0.013***	0.010***	0.007**	0.040***	0.041***	0.040***
	(0.007)	(0.007)	(0.008)	(0.003)	(0.003)	(0.003)	(0.012)	(0.013)	(0.015)
Interaction	0.105***	0.109***	0.103***	0.000	0.001	0.006	0.050	0.067**	0.123***
	(0.016)	(0.018)	(0.021)	(0.006)	(0.007)	(0.008)	(0.033)	(0.034)	(0.034)
Observations	345944	291430	240227	327987	279016	230758	77815	71910	58286
Regions	1787	1555	1340	1755	1530	1315	668	617	545
Adj. R-Square	0.16	0.16	0.17	0.38	0.39	0.42	0.22	0.23	0.25
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region F.E.	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (col. (1)-(3)), religiosity dummy (col. (4)-(6)) and gender-egalitarian index (col. (7)-(9)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. The sample of countries in analysis is based on countries area of land in KM^2: above 50000 KM^2 (col. (1), (4) and (7)), above 100000 KM^2 (col. (2), (5) and (8)) and above 200000 KM^2 (col. (3), (6) and (9)).

Table A14: Connectedness and Culture - Multiple Connections

	(1) OLS 2009-2012	(2) OLS 2009-2012	(3) LPM 2009-2012	(4) LPM 2009-2012	(5) OLS 2009-2011	(6) OLS 2009-2011
	Social Behavior		*Religiosity*		*Gender-Egalitarian*	
$\sum$ *Network*	0.154***		0.011***		0.028***	
	(0.005)		(0.002)		(0.009)	
$Network_1$		0.178***		0.014***		0.047***
		(0.006)		(0.002)		(0.012)
$Network_2$		0.080***		-0.024**		-0.038
		(0.014)		(0.010)		(0.042)
$Network_3$		0.129***		-0.010		0.013
		(0.022)		(0.014)		(0.061)
$Interaction_1$	0.063***	0.065***	-0.005	-0.007	0.048*	0.049*
	(0.014)	(0.014)	(0.006)	(0.006)	(0.029)	(0.029)
$Interaction_2$	-0.011	0.042*	-0.017**	0.030**	0.031	0.077
	(0.020)	(0.021)	(0.008)	(0.014)	(0.044)	(0.050)
$Interaction_3$	-0.019	0.034	-0.006	0.029	-0.043	-0.010
	(0.031)	(0.033)	(0.011)	(0.019)	(0.058)	(0.070)
Observations	412136	412136	392653	392653	90316	90316
Regions	2097	2097	2065	2065	736	736
Adj. R-Square	0.16	0.16	0.38	0.38	0.22	0.22
Individual Controls	✓	✓	✓	✓	✓	✓
Region F.E.	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (col. (1)-(2)), religiosity dummy (col. (3)-(4)) and gender-egalitarian index (col. (5)-(6)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. The table report the coefficients associated to the total number of connections (max 3), to the first, second and third connection and their interaction term with the average culture of connection's country of residence.

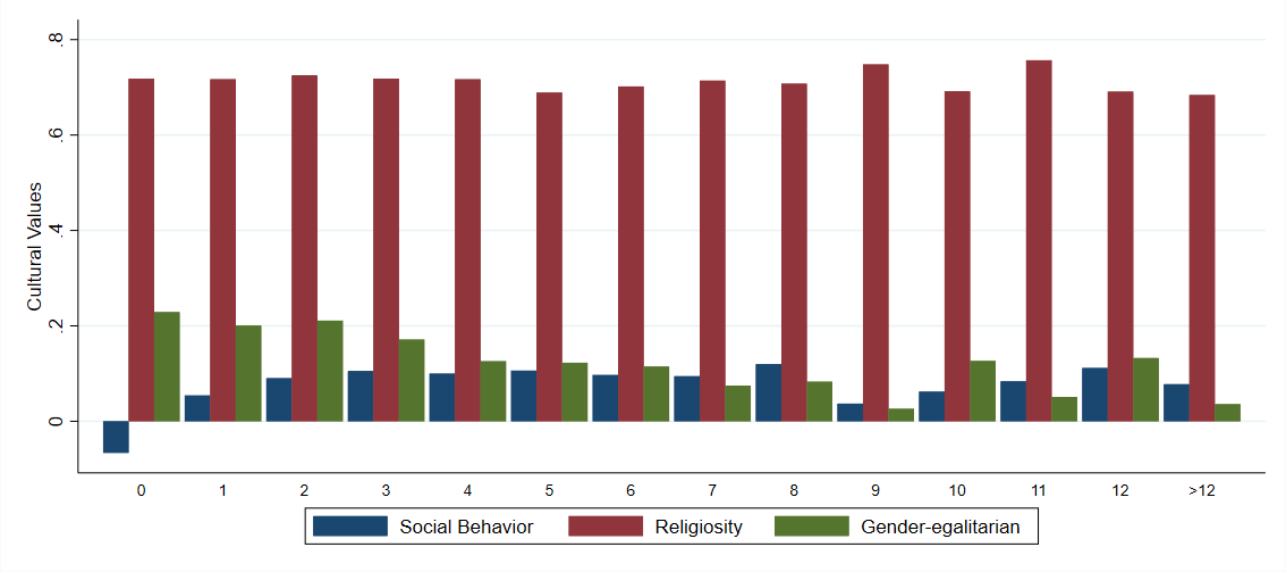
6.8.6 Controlling for General Connectedness

Table A15: Controlling for General Connectedness - Subsamples

	(1) OLS 2009-2012	(2) OLS 2009-2012	(3) LPM 2009-2012	(4) LPM 2009-2012	(5) OLS 2009-2011	(6) OLS 2009-2011
	Social Behavior		*Religiosity*		*Gender-Egalitarian*	
Rel. Connection	No	Yes	No	Yes	No	Yes
Network	0.191***	0.187***	0.012**	0.012***	0.051*	0.028**
	(0.014)	(0.006)	(0.005)	(0.003)	(0.028)	(0.014)
Interaction	0.061	0.104***	0.004	-0.005	0.015	0.074**
	(0.042)	(0.014)	(0.014)	(0.007)	(0.084)	(0.037)
Observations	76353	301754	72430	289569	19498	56275
Regions	1956	2084	1919	2045	653	707
Adj. R-Square	0.17	0.17	0.35	0.38	0.25	0.23
Region F.E.	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. Columns (1), (3) and (5) show the estimates from the specification presented in equation (6.1) while columns (2), (4) and (6) show the estimates from the specification presented in equation (6.2). The dependent variable is: social-behavior index (col. (1)-(2)), religiosity dummy (col. (3)-(4)) and gender-egalitarian index (col. (5)-(6)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. The analysis is performed on subsamples based on whether individuals have a reliable connection in general: individuals without a connection (col. (1), (3) and (5)) and individuals with a connection (col. (2), (4) and (6)).

Figure A3: Social Activities and Cultural Traits



Note: authors' calculations on Gallup World Poll Data. The Figure plots the average cultural trait associated to groups of individuals who spent from 0 to more than 12 hours in social activities with relatives and friends the day before the survey. Each group of individuals is represented on the x-axis.

6.8.7 Selection on Unobservables

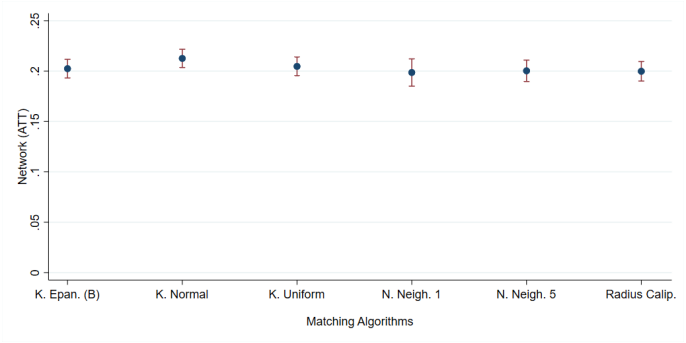
Table A16: Selection on Unobservables - No interaction

	(1)	(2)	(3)	(4)	(5)
		$R_{max} = 1.3\tilde{R}$		$R_{max} = 2\tilde{R}$	
	Benchmark	δ	*Id. Set*	δ	*Id. Set*
Panel A - Social Behavior (OLS)					
Network	0.205***	6.452	[0.185; 0.205]	1.959	[0.122; 0.205]
	(0.006)				
Adj. R-Square ($\tilde{R}$)	0.16				
Panel B - Religiosity (LPM)					
Network	0.012***	-4.372	[0.008; 0.012]	-1.222	[0.002; 0.012]
	(0.002)				
Adj. R-Square ($\tilde{R}$)	0.38				
Panel C - Gender-Egalitarian (OLS)					
Network	0.054***	-4.654	[0.041; 0.054]	-1.265	[0.009; 0.054]
	(0.011)				
Adj. R-Square ($\tilde{R}$)	0.22				

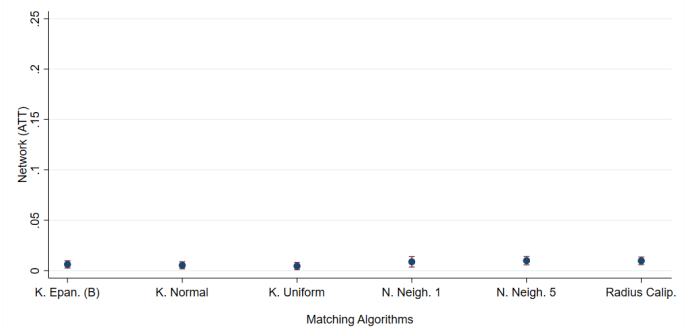
Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level in column (1). * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (Panel A), religiosity dummy (Panel B) and gender-egalitarian index (Panel C). Column (1) show the estimates of the model presented in equation (6.1). Columns (2) and (4) shows the value of selection on unobservables (δ) which produces $\beta = 0$ given the value of R_{max}. Columns (3) and (5) shows the identified set of the estimated $\hat{\beta}$ when $\delta = 0$ (no bias-adjustment) and $\tilde{\beta}$ when $\delta = 1/-1$ (observables as important as unobservables) given the value of R_{max}. Columns (2) and (3) shows the results for the suggested level of R_{max} by Oster (2019).

6.8.8 Matching Results

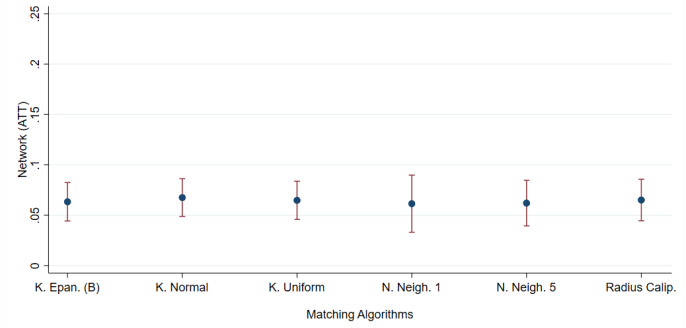
Figure A4: Average Effect of Connection - Different Matching Algorithms



(a) Social Behavior



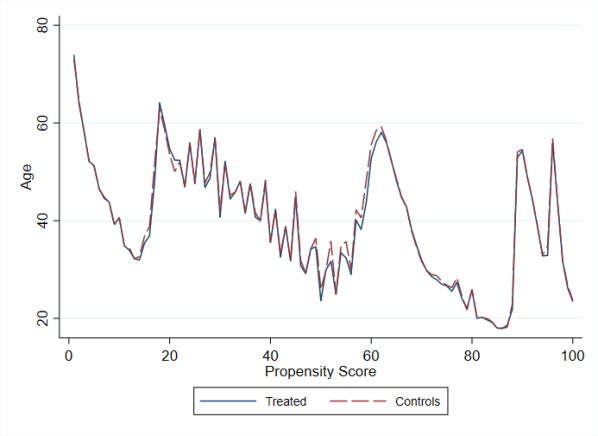
(b) Religiosity



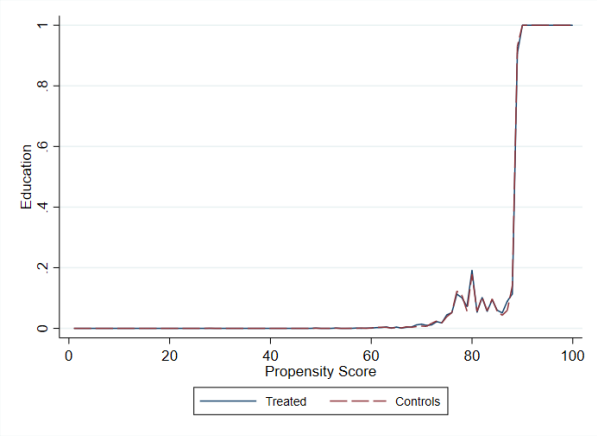
(c) Gender-Egalitarian

Note: authors' calculations on Gallup World Poll Data. The Figure plots the average effect of having a connection abroad after a propensity score matching on three different cultural traits: Social Behavior (Figure a), Religiosity (Figure b) and Gender-Egalitarian (Figure c) and the interval of confidence at 99% level. Each figure shows the results using the propensity score computed in column (2) in Table A1 and different matching algorithms: Kernel (Epanechnikov, Normal and Uniform) matching, Nearest Neighbour Matching (1 and 5 individuals with replacement) and Radius Matching.

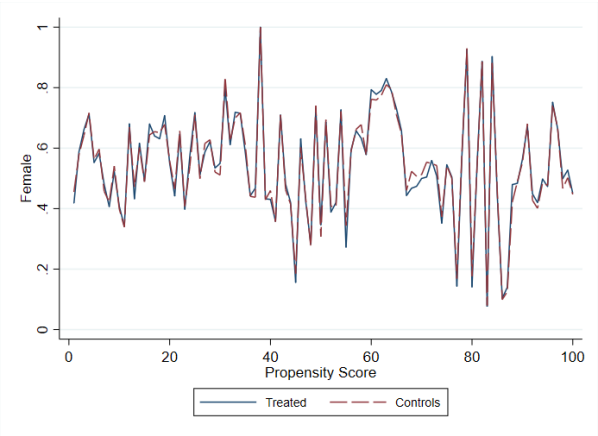
Figure A5: Distribution Covariates after Matching - Social Behavior



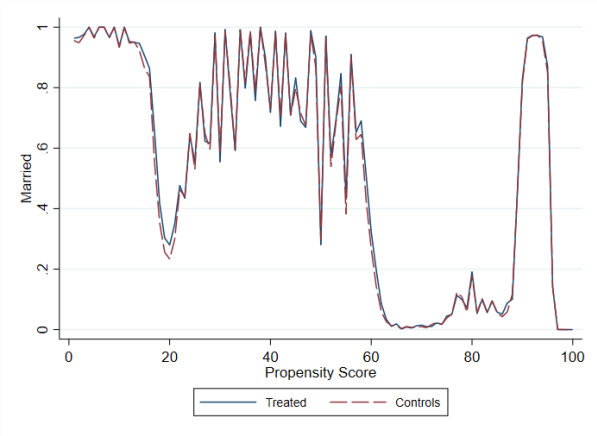
(a) Age



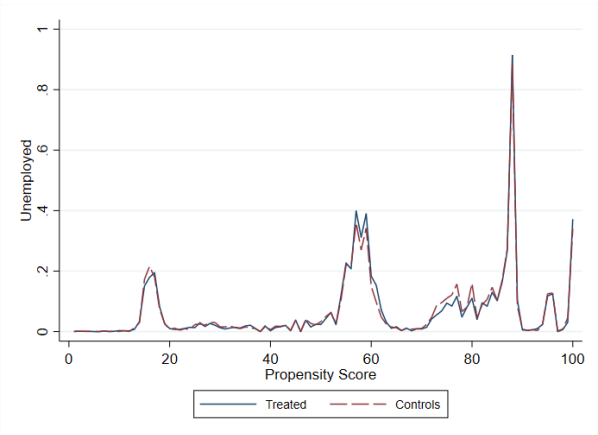
(b) Education



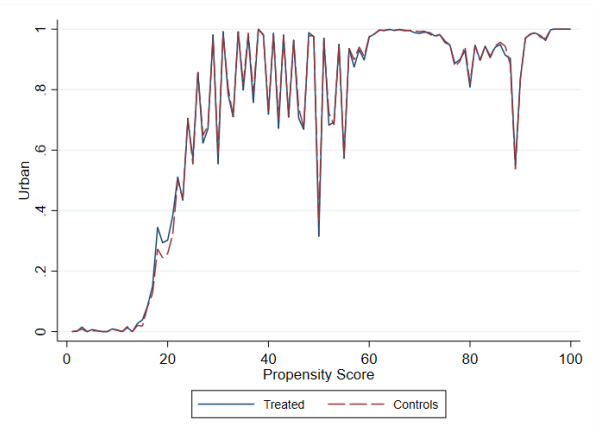
(c) Gender



(d) Marital



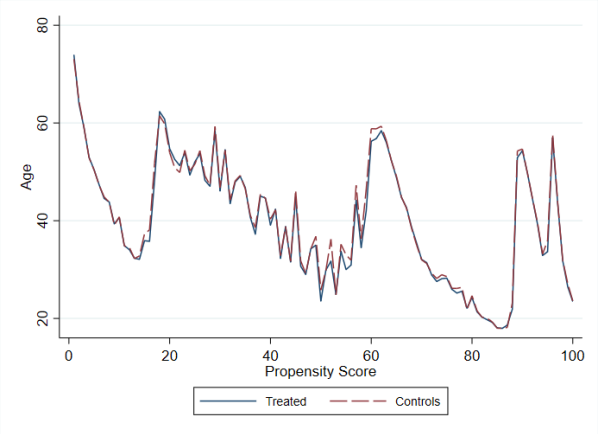
(e) Unemployment



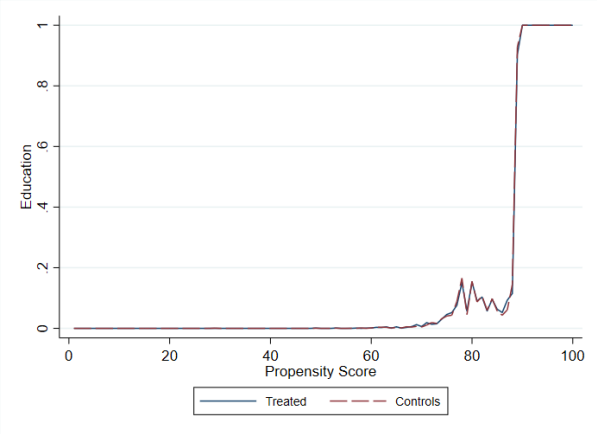
(f) Urban

Note: authors' calculations on Gallup World Poll Data. The Figure plots the distribution of covariates by centiles of propensity score.

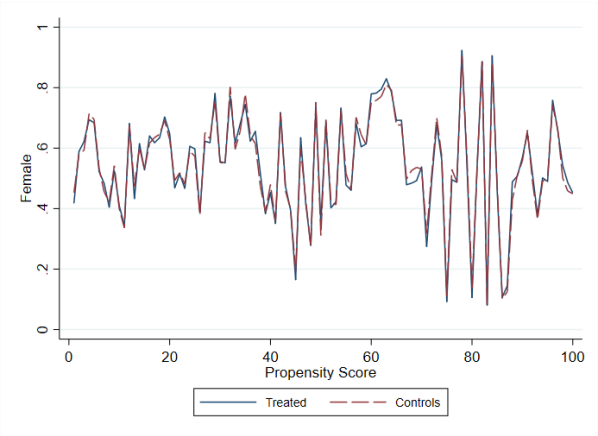
Figure A6: Distribution Covariates after Matching - Religiosity



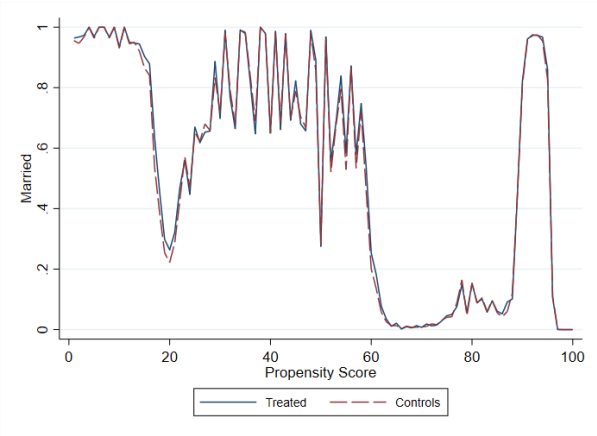
(a) Age



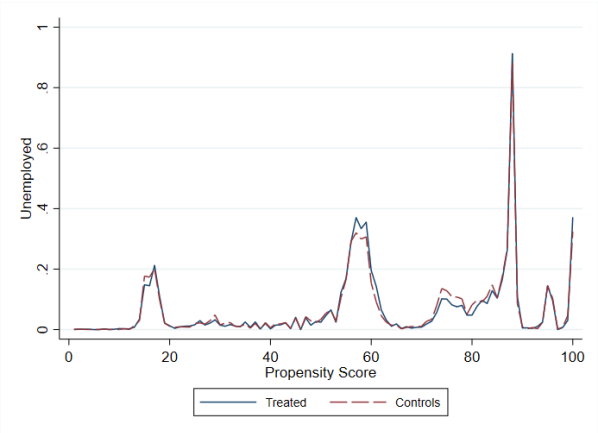
(b) Education



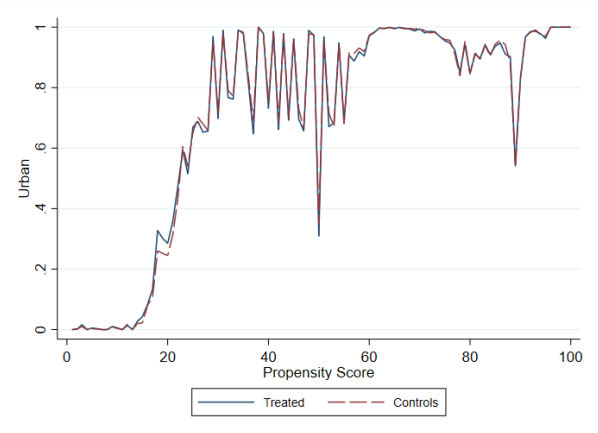
(c) Gender



(d) Marital



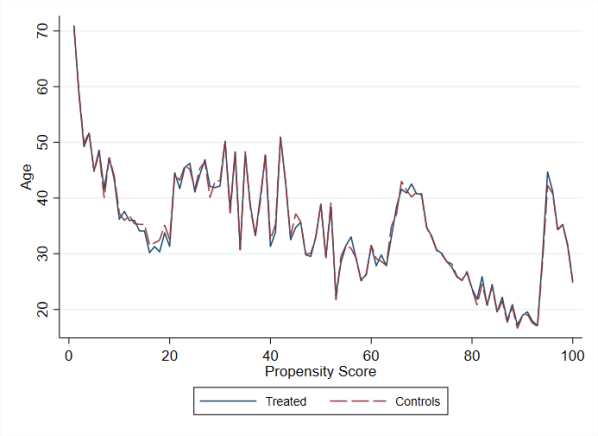
(e) Unemployment



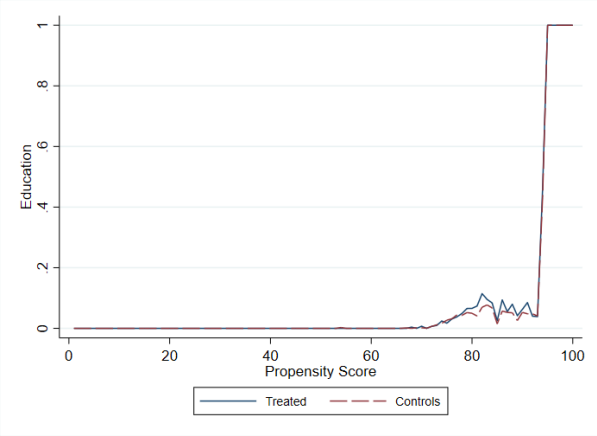
(f) Urban

Note: authors' calculations on Gallup World Poll Data. The Figure plots the distribution of covariates by centiles of propensity score.

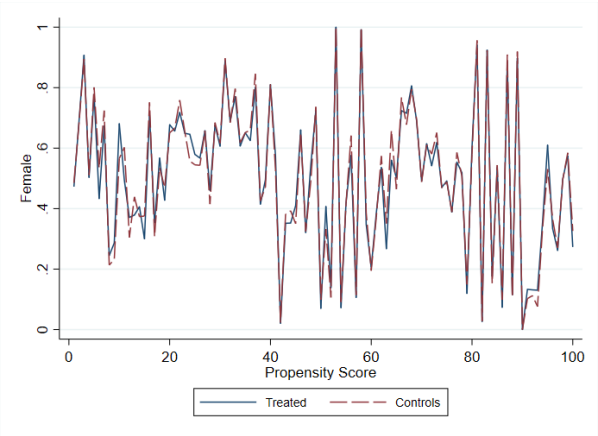
Figure A7: Distribution Covariates after Matching - Gender-Egalitarian



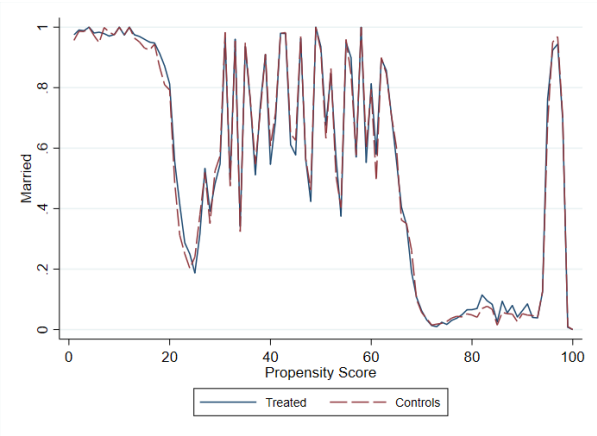
(a) Age



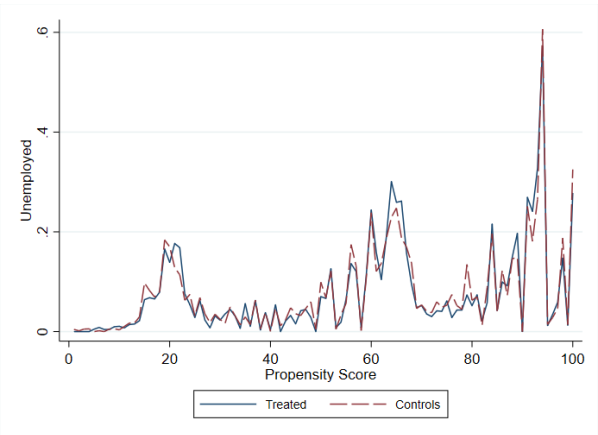
(b) Education



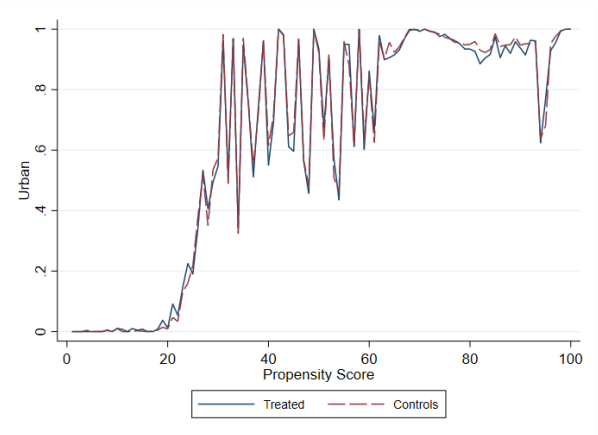
(c) Gender



(d) Marital



(e) Unemployment



(f) Urban

Note: authors' calculations on Gallup World Poll Data. The Figure plots the distribution of covariates by centiles of propensity score.

6.8.9 Cultural Convergence or Divergence

Table A17: Cultural Convergence or Divergence - Regional subsamples by regional culture (Median)

	(1) OLS 2009-2012	(2) OLS 2009-2012	(3) LPM 2009-2012	(4) LPM 2009-2012	(5) OLS 2009-2011	(6) OLS 2009-2011
	Social Behavior		*Religiosity*		*Gender-Egalitarian*	
Median Trait	Below	Above	Below	Above	Below	Above
Panel A - Average						
Network	0.184***	0.209***	0.027***	0.005**	0.067***	0.010
	(0.007)	(0.010)	(0.004)	(0.002)	(0.015)	(0.016)
Interaction	0.086***	0.102***	-0.022*	0.003	0.054	0.046
	(0.019)	(0.019)	(0.013)	(0.004)	(0.037)	(0.047)
Observations	246093	165274	176830	215071	52764	37475
Regions	1143	954	999	1066	366	370
Adj. R-Square	0.08	0.09	0.21	0.05	0.18	0.08
Panel B - Distance						
Network	0.181***	0.223***	0.023***	0.007***	0.062***	0.013
	(0.008)	(0.010)	(0.004)	(0.002)	(0.017)	(0.016)
Distance	0.051***	0.080***	0.082***	0.015*	0.048*	0.072
	(0.015)	(0.021)	(0.016)	(0.008)	(0.027)	(0.058)
Observations	246093	165274	176830	215071	52764	37475
Regions	1143	954	999	1066	366	370
Adj. R-Square	0.08	0.09	0.21	0.05	0.18	0.08
Individual Controls	✓	✓	✓	✓	✓	✓
Region F.E.	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (col. (1)-(2)), religiosity dummy (col. (3)-(4)) and gender-egalitarian index (col. (5)-(6)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. The sample of regions is splitted by cultural traits median, respectively: below the median (col. (1),(3),(5)) and above the median (col. (2),(4),(6)). Panel A shows the estimates using the specification presented in equation (6.2), while Panel B replace the average culture of connection's country of residence with a measure of cultural distance.

Table A18: Cultural Convergence or Divergence - Regional subsamples by regional culture (Quartiles)

	(1) OLS 2009-2012	(2) OLS 2009-2012	(3) OLS 2009-2012	(4) OLS 2009-2012	(5) LPM 2009-2012	(6) LPM 2009-2012	(7) LPM 2009-2012	(8) LPM 2009-2012	(9) OLS 2009-2011	(10) OLS 2009-2011	(11) OLS 2009-2011	(12) OLS 2009-2011
	Social Behavior				*Religiosity*				*Gender-Egalitarian*			
Trait Quartile	1^{st}	2^{nd}	3^{rd}	4^{th}	1^{st}	2^{nd}	3^{rd}	4^{th}	1^{st}	2^{nd}	3^{rd}	4^{th}
Panel A - Average												
Network	0.169***	0.198***	0.208***	0.211***	0.031***	0.025***	0.006	0.003*	0.081***	0.051**	0.008	0.015
	(0.009)	(0.010)	(0.015)	(0.014)	(0.006)	(0.006)	(0.004)	(0.002)	(0.023)	(0.021)	(0.025)	(0.020)
Interaction	0.061**	0.104***	0.105***	0.098***	0.019	-0.059***	0.006	0.001	0.053	0.065	0.053	0.050
	(0.029)	(0.025)	(0.032)	(0.024)	(0.019)	(0.017)	(0.009)	(0.003)	(0.051)	(0.051)	(0.078)	(0.046)
Observations	125706	120387	81322	83952	84757	92073	109802	105269	25017	27746	17072	20403
Regions	624	519	373	581	465	534	546	520	167	199	146	224
Adj. R-Square	0.06	0.04	0.04	0.08	0.09	0.10	0.03	0.01	0.18	0.08	0.05	0.05
Panel B - Distance												
Network	0.166***	0.195***	0.212***	0.231***	0.030***	0.018***	0.010***	0.004**	0.072***	0.051**	0.009	0.021
	(0.010)	(0.011)	(0.015)	(0.013)	(0.006)	(0.006)	(0.004)	(0.002)	(0.026)	(0.021)	(0.025)	(0.019)
Distance	0.031	0.086***	0.099***	0.080***	0.075***	0.064***	0.025	0.007	0.045	0.056	0.076	0.072
	(0.020)	(0.023)	(0.035)	(0.025)	(0.027)	(0.024)	(0.017)	(0.006)	(0.035)	(0.048)	(0.091)	(0.059)
Observations	125706	120387	81322	83952	84757	92073	109802	105269	25017	27746	17072	20403
Regions	624	519	373	581	465	534	546	520	167	199	146	224
Adj. R-Square	0.06	0.04	0.04	0.08	0.09	0.10	0.03	0.01	0.18	0.08	0.05	0.05
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region F.E.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (col. (1)-(4)), religiosity dummy (col. (5)-(8)) and gender-egalitarian index (col. (9)-(12)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. The sample of regions is splitted by cultural traits quartiles, respectively: first quartile (col. (1),(5),(9)), second quartile (col. (2),(6),(10)), third quartile (col. (3),(7),(11)) and fourth quartile (col. (4),(8),(12)). Panel A shows the estimates using the specification presented in equation (6.2), while Panel B replace the average culture of connection's country of residence with a measure of cultural distance.

Table A19: Cultural Convergence or Divergence - Broad Continents

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Broad Geography	Europe	Former SU	Asia	Latina America	MENA	Sub-Saharan	North America	Oceania
Panel A - Social Behavior (OLS)								
Network	0.194***	0.162***	0.208***	0.247***	0.157***	0.196***	0.175***	0.096
	(0.011)	(0.013)	(0.019)	(0.016)	(0.015)	(0.013)	(0.050)	(0.065)
Interaction	0.072***	0.103**	0.098**	0.034	0.100**	0.181***	0.215***	0.198**
	(0.023)	(0.043)	(0.040)	(0.027)	(0.040)	(0.049)	(0.078)	(0.092)
Observations	67754	54473	79680	56186	56549	89400	4569	2756
Regions	406	237	339	377	258	395	61	24
Adj. R-Square	0.21	0.13	0.19	0.09	0.13	0.13	0.09	0.05
Panel B - Religiosity (LPM)								
Network	0.028***	0.017**	0.020***	0.016***	0.003	0.008**	0.017	0.047**
	(0.007)	(0.008)	(0.006)	(0.006)	(0.005)	(0.003)	(0.022)	(0.018)
Interaction	-0.019	0.028	-0.006	-0.009	-0.024*	0.002	-0.015	0.082
	(0.020)	(0.024)	(0.015)	(0.014)	(0.014)	(0.005)	(0.061)	(0.068)
Observations	67754	54473	76217	56186	39511	89798	5206	2756
Regions	406	237	339	377	226	395	61	24
Adj. R-Square	0.21	0.24	0.54	0.16	0.14	0.06	0.16	0.04
Panel C - Gender-Egalitarian (OLS)								
Network	-	0.073***	0.005	-	0.080***	0.029*	-	-
	-	(0.025)	(0.038)	-	(0.026)	(0.015)	-	-
Interaction	-	-0.046	0.139**	-	0.012	0.076***	-	-
	-	(0.118)	(0.063)	-	(0.084)	(0.033)	-	-
Observations	-	8540	11374	-	21281	49044	-	-
Regions	-	54	89	-	228	365	-	-
Adj. R-Square	-	0.16	0.35	-	0.18	0.22	-	-
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓
Region F.E.	✓	✓	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (Panel A), religiosity dummy (Panel B) and gender-egalitarian index (Panel C). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. Each column estimates the benchmark model in equation (6.2) on a different subset of countries of individuals based on their broad geographical area: Europe (col. (1)), Former Soviet Union (col. (2)), Asia (col. (3)), Latina America (col. (4)), Middle East and Noth Africa (col. (5)), Sub-Saharan Africa (col. (6)), North America (col. (7)) and Oceania (col. (8)). For Panel C, estimates associate to Europe, Latina America, North America and Oceania are not available since GWP does not ask questions about gender-egalitarian attitudes in those regions.

Table A20: CF by Trait and F_{st} by Cleavage

	(1) # Regions	(2) Mean	(3) SD	(4) Min	(5) Max
Panel A - Social Behavior					
CF	2197	0.677	0.124	0	0.894
F_{st} Connectedness	2197	0.025	0.038	0	0.556
F_{st} Education	2197	0.016	0.028	0	0.645
F_{st} Unemployed	2197	0.016	0.030	0	0.540
Panel B - Religiosity					
CF	2041	0.255	0.170	0	0.499
F_{st} Connectedness	2041	0.021	0.054	0	1
F_{st} Education	2041	0.017	0.055	0	1
F_{st} Unemployed	2041	0.016	0.045	0	1
Panel C - Gender-egalitarian					
CF	805	0.562	0.188	0	0.857
F_{st} Connectedness	805	0.031	0.055	0	1
F_{st} Education	805	0.009	0.026	0	0.321
F_{st} Unemployed	805	0.021	0.041	0	0.421

Note: authors' calculations on Gallup World poll data. Each panel presents the results related to: social behavior (A), religiosity (B) and gender-egalitarian attitudes (C). Each panel shows the average regional total heterogeneity (CF) and the share of the total heterogeneity which is not attributable to withing-groups heterogeneity. The three groups-cleavages are based on: connectedness, education and employment status.

6.8.10 Connectedness Diffusion

Table A21: Country of origin heterogeneity - Network presence (Median)

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	LPM	LPM	OLS	OLS
	2009-2012	2009-2012	2009-2012	2009-2012	2009-2011	2009-2011
	Social Behavior		*Religiosity*		*Gender-Egalitarian*	
Median Network	Below	Above	Below	Above	Below	Above
Network	0.213***	0.185***	0.011***	0.016***	0.023	0.067***
	(0.010)	(0.007)	(0.004)	(0.003)	(0.016)	(0.016)
Interaction	0.131***	0.090***	-0.012	-0.004	0.108**	0.017
	(0.024)	(0.016)	(0.008)	(0.008)	(0.054)	(0.035)
Observations	228655	182712	215379	176522	54616	35623
Regions	1201	896	1189	876	526	210
Adj. R-Square	0.17	0.15	0.40	0.35	0.22	0.22
Individual Controls	✓	✓	✓	✓	✓	✓
Region F.E.	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (col. (1)-(2)), religiosity dummy (col. (3)-(4)) and gender-egalitarian index (col. (5)-(6)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. The sample of regions is splitted by median based on the regional share of connected individuals: below median (col. (1),(3),(5)) and above median (col. (2),(4),(6)).

Table A22: Country of origin heterogeneity - Network presence (Quartiles)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	OLS	OLS	OLS	OLS	LPM	LPM	LPM	LPM	OLS	OLS	OLS	OLS
	2009-2012	2009-2012	2009-2012	2009-2012	2009-2012	2009-2012	2009-2012	2009-2012	2009-2011	2009-2011	2009-2011	2009-2011
	Social Behavior				*Religiosity*				*Gender-Egalitarian*			
Network Quartile	1^{st}	2^{nd}	3^{rd}	4^{th}	1^{st}	2^{nd}	3^{rd}	4^{th}	1^{st}	2^{nd}	3^{rd}	4^{th}
Network	0.214***	0.212***	0.187***	0.181***	0.015**	0.008*	0.017***	0.015***	0.012	0.033	0.058**	0.070***
	(0.014)	(0.013)	(0.010)	(0.011)	(0.006)	(0.004)	(0.005)	(0.004)	(0.025)	(0.021)	(0.026)	(0.019)
Interaction	0.128***	0.128***	0.091***	0.091***	-0.021	-0.002	0.004	-0.010	0.207**	0.068	-0.013	0.034
	(0.039)	(0.029)	(0.028)	(0.020)	(0.016)	(0.009)	(0.011)	(0.010)	(0.086)	(0.066)	(0.055)	(0.046)
Observations	121178	107477	79538	103170	115910	99469	75688	100832	29440	25176	16310	19313
Regions	658	543	360	535	653	536	352	524	307	219	105	105
Adj. R-Square	0.18	0.16	0.14	0.16	0.45	0.33	0.33	0.37	0.23	0.21	0.18	0.26
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region F.E.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (col. (1)-(4)), religiosity dummy (col. (5)-(8)) and gender-egalitarian index (col. (9)-(12)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. The sample of regions is splitted by quartiles based on the regional share of connected individuals: first quartile (col. (1),(5),(9)), second quartile (col. (2),(6),(10)), third quartile (col. (3),(7),(11)) and fourth quartile (col. (4),(8),(12)).

6.8.11 Individual Heterogeneity

Table A23: Individuals heterogeneity - Natives and Migrants

	(1)	(2)	(3)	(4)	(5)	(6)
Cultural Trait	*Social Behavior*		*Religiosity*		*Gender-Egalitarian*	
Migration Status	Natives	Migrants	Natives	Migrants	Natives	Migrants
Network	0.196***	0.179***	0.013***	-0.031***	0.046***	0.077
	(0.006)	(0.022)	(0.002)	(0.011)	(0.012)	(0.052)
Interaction	0.100***	0.065	-0.008	0.083***	0.055*	0.255**
	(0.013)	(0.044)	(0.006)	(0.024)	(0.030)	(0.100)
Observations	411367	17107	391901	15222	90239	3207
Regions	2097	1002	2065	985	736	239
Adj. R-Square	0.16	0.21	0.38	0.26	0.22	0.18
Individual Controls	✓	✓	✓	✓	✓	✓
Region F.E.	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (col. (1) and (2)), religiosity dummy (col. (3) and (4)) and gender-egalitarian index (col. (5) and (6)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. The analysis is performed on subsamples based on individual migration status: natives (col. (1), (3) and (5)) and migrants (col. (2), (4) and (6)).

Table A24: Individuals heterogeneity - Education, same sample of regions

	(1)	(2)	(3)	(4)	(5)	(6)
Cultural Trait	*Social Behavior*		*Religiosity*		*Gender-Egalitarian*	
Education Level	LS	HS	LS	HS	LS	HS
Network	0.193***	0.196***	0.016***	0.012*	0.052***	0.042
	(0.007)	(0.012)	(0.003)	(0.007)	(0.014)	(0.027)
Interaction	0.115***	0.035	-0.007	-0.012	0.038	0.022
	(0.016)	(0.029)	(0.006)	(0.017)	(0.034)	(0.086)
Observations	326943	53356	309102	51228	63933	6779
Regions	1761	1762	1730	1731	406	406
Adj. R-Square	0.15	0.18	0.37	0.31	0.20	0.24
Individual Controls	✓	✓	✓	✓	✓	✓
Region F.E.	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (col. (1) and (2)), religiosity dummy (col. (3) and (4)) and gender-egalitarian index (col. (5) and (6)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. The analysis is performed on subsamples based on individual education level: less then nine years of education (col. (1), (3) and (5)) and more then nine years of education (col. (2), (4) and (6)).

6.8.12 Differential effects based on connection location

Table A25: Cultural Convergence or Divergence - Country subsamples by connection culture (Median)

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	LPM	LPM	OLS	OLS
	2009-2012	2009-2012	2009-2012	2009-2012	2009-2011	2009-2011
	Social Behavior		*Religiosity*		*Gender-Egalitarian*	
Median Trait	Below	Above	Below	Above	Below	Above
Network	0.198***	0.195***	0.014***	0.017***	0.048***	0.049***
	(0.007)	(0.007)	(0.002)	(0.002)	(0.013)	(0.014)
Interaction	0.127***	0.099***	-0.013*	-0.005	0.050	0.050
	(0.045)	(0.016)	(0.008)	(0.006)	(0.120)	(0.033)
Observations	379904	398736	384849	355367	84670	87391
Regions	2095	2097	2064	2065	736	736
Adj. R-Square	0.16	0.16	0.38	0.38	0.22	0.22
Individual Controls	✓	✓	✓	✓	✓	✓
Region F.E.	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (col. (1)-(3)), religiosity dummy (col. (4)-(6)) and gender-egalitarian index (col. (7)-(9)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. The sample of individuals is splitted by average culture of connection's country of location median, respectively: below the median (col. (1),(3),(5)) and above the median (col. (2),(4),(6)).

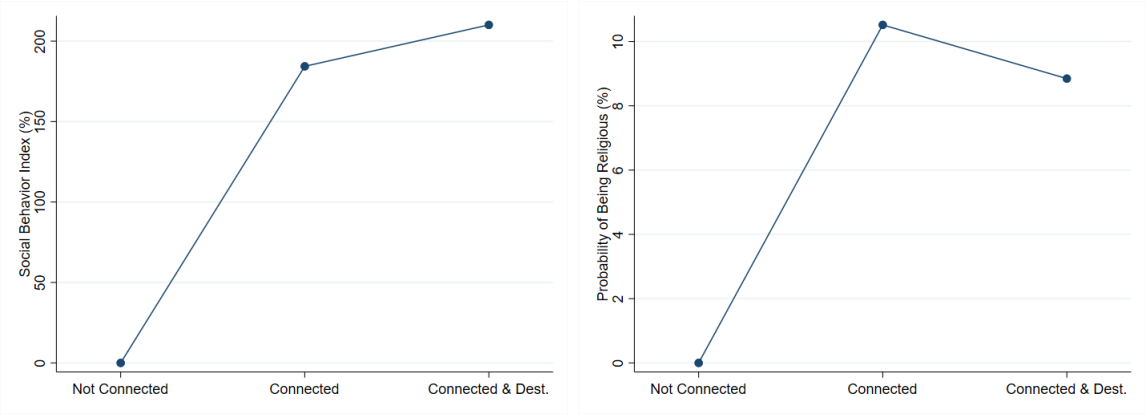
Table A26: Cultural Convergence or Divergence - Country subsamples by connection culture (Terciles)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	OLS	OLS	OLS	LPM	LPM	LPM	OLS	OLS	OLS
	2009-2012	2009-2012	2009-2012	2009-2012	2009-2012	2009-2012	2009-2011	2009-2011	2009-2011
	Social Behavior			*Religiosity*			*Gender-Egalitarian*		
Tertile Trait	1^{st}	2^{nd}	3^{rd}	1^{st}	2^{nd}	3^{rd}	1^{st}	2^{nd}	3^{rd}
Network	0.197***	0.196***	0.194***	0.014***	0.017***	0.017***	0.048***	0.049***	0.053***
	(0.007)	(0.007)	(0.007)	(0.002)	(0.002)	(0.002)	(0.013)	(0.014)	(0.014)
Interaction	0.127***	-0.430	0.102***	-0.012	-0.003	-0.009	0.050	-0.478	0.047
	(0.045)	(0.349)	(0.017)	(0.008)	(0.011)	(0.006)	(0.120)	(0.570)	(0.033)
Observations	376764	377826	391323	384079	350647	353805	84670	82448	86765
Regions	2095	2095	2097	2064	2064	2065	736	736	736
Adj. R-Square	0.16	0.16	0.16	0.38	0.38	0.38	0.22	0.22	0.22
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region F.E.	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country-year F.E.	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: authors' calculations on Gallup World poll data. Standard errors are clustered at regional level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is: social-behavior index (col. (1)-(3)), religiosity dummy (col. (4)-(6)) and gender-egalitarian index (col. (7)-(9)). The set of individual controls includes a dummies for education, gender, marital status, having children, living in an urban area, being unemployed and continuous variables for age, family size and income. The sample of individuals is splitted by average culture of connection's country of location terciles, respectively: first tercile (col. (1),(4),(7)), second tercile (col. (2),(5),(8)) and third tercile (col. (3),(6),(9)).

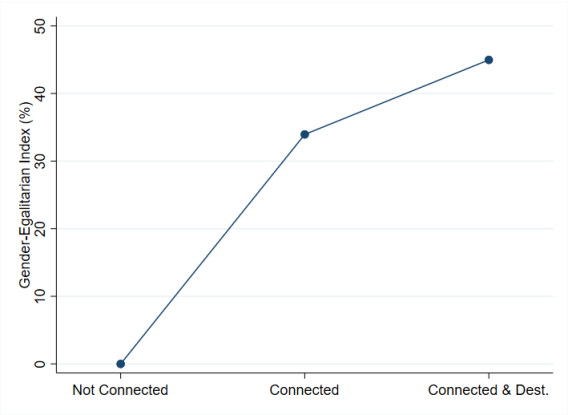
6.8.13 How much does it count?

Figure A8: Relative deviation from the No-connection Scenario - Individual Level



(a) Social Behavior

(b) Religiosity



(c) Gender-egalitarian

Note: authors' calculations on Gallup World Poll Data. The Figure plots the simulated cultural effect of connectedness on each cultural trait. Figures (a), (b) and (c) shows the relative deviation each cultural trait between individuals who has no connection and who hasa connection abroad. The effects are simulated using coefficients from Table 6.3

Table A27: Simulations: country values

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
		Social Behavior				Religiosity				Gender-Egalitarian			
Country	Network	Actual	No Con.	Con.	Inter.	Actual	No Con.	Con.	Inter.	Actual	No Con.	Con.	Inter.
Afghanistan	0.18	0.17	0.12	0.16	0.16	0.97	0.97	0.97	0.97	-0.90	-0.90	-0.89	-0.89
Albania	0.46	-0.53	-0.55	-0.46	-0.46	0.42	0.43	0.43	0.43	0.32	0.22	0.25	0.27
Algeria	0.39	-0.40	-0.52	-0.44	-0.43	0.93	0.94	0.94	0.94	0.10	-0.01	0.02	0.04
Angola	0.32	0.04	-0.08	-0.01	-0.01	0.84	0.81	0.81	0.81	-0.65	-0.65	-0.63	-0.63
Argentina	0.25	-0.12	-0.20	-0.15	-0.14	0.63	0.64	0.64	0.64	.	.	.	.
Armenia	0.41	-0.34	-0.36	-0.28	-0.28	0.74	0.74	0.74	0.74	.	.	.	.
Australia	0.50	0.84	0.83	0.93	0.96	0.32	0.30	0.30	0.30	.	.	.	.
Austria	0.41	0.36	0.27	0.35	0.36	0.49	0.50	0.51	0.51	.	.	.	.
Azerbaijan	0.16	-0.05	-0.08	-0.04	-0.05	0.45	0.44	0.45	0.45	0.17	0.15	0.16	0.16
Bahrain	0.27	0.06	-0.14	-0.09	-0.09	0.95	0.95	0.95	0.95	-0.05	0.04	0.06	0.06

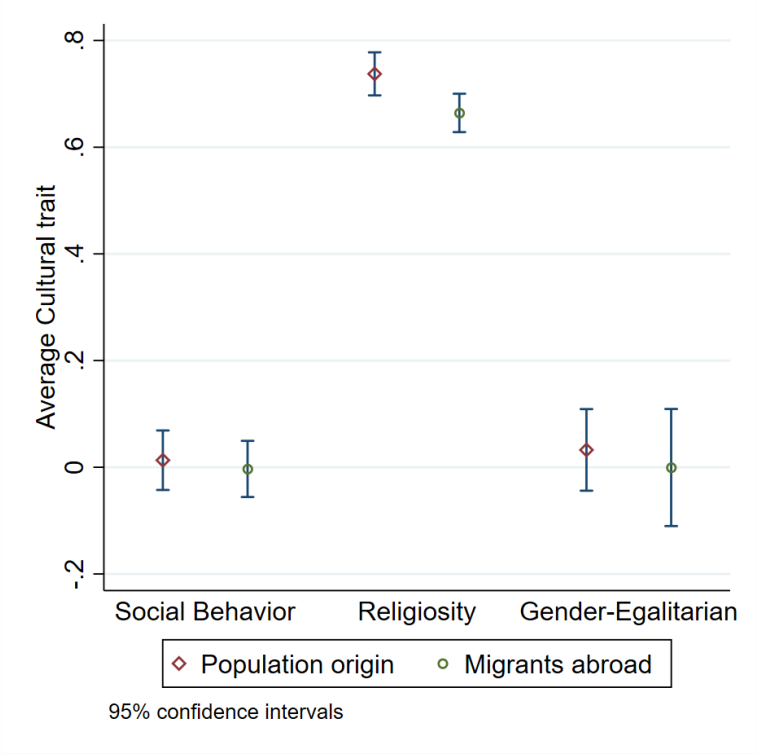
Bangladesh	0.25	-0.26	-0.30	-0.25	-0.25	0.99	0.99	0.99	0.99	0.06	0.07	0.08	0.08
Belarus	0.29	-0.13	-0.21	-0.15	-0.15	0.38	0.37	0.37	0.37	.	.	.	.
Belgium	0.37	0.09	0.03	0.11	0.11	0.35	0.31	0.31	0.31	.	.	.	.
Benin	0.25	-0.46	-0.52	-0.47	-0.47	0.83	0.83	0.83	0.83	-0.06	-0.07	-0.05	-0.05
Bolivia	0.57	-0.01	-0.17	-0.05	-0.05	0.87	0.86	0.87	0.87	.	.	.	.
Bosnia & Herzeg.	0.36	-0.38	-0.45	-0.37	-0.36	0.71	0.69	0.70	0.70	0.56	0.57	0.60	0.61
Botswana	0.26	-0.14	-0.22	-0.17	-0.17	0.87	0.85	0.85	0.85	0.37	0.38	0.40	0.40
Brazil	0.14	-0.12	-0.16	-0.13	-0.13	0.88	0.89	0.89	0.89	.	.	.	.
Bulgaria	0.31	-0.46	-0.52	-0.46	-0.45	0.41	0.39	0.39	0.39	.	.	.	.
Burkina Faso	0.45	-0.37	-0.41	-0.32	-0.33	0.94	0.93	0.94	0.94	0.14	0.15	0.18	0.18
Burundi	0.08	-0.59	-0.60	-0.59	-0.59	0.96	0.97	0.97	0.97	-0.07	-0.07	-0.07	-0.06
Cambodia	0.15	-0.10	-0.14	-0.11	-0.10	0.94	0.94	0.94	0.94	.	.	.	.
Cameroon	0.32	-0.04	-0.14	-0.07	-0.07	0.95	0.95	0.95	0.95	-0.06	-0.06	-0.04	-0.03
Canada	0.40	0.79	0.67	0.75	0.77	0.41	0.40	0.40	0.40	.	.	.	.
Central African Rep.	0.23	-0.18	-0.23	-0.19	-0.19	0.98	0.98	0.98	0.98	0.14	0.13	0.14	0.15
Chad	0.29	-0.20	-0.26	-0.20	-0.20	0.92	0.91	0.92	0.92	0.04	0.02	0.03	0.04
Chile	0.30	0.21	0.12	0.18	0.19	0.69	0.70	0.70	0.70	.	.	.	.
China	0.03	-0.51	-0.53	-0.52	-0.52	0.15	0.15	0.15	0.15	.	.	.	.
Colombia	0.48	0.18	0.09	0.18	0.20	0.86	0.88	0.88	0.88	.	.	.	.
Comoros	0.53	-0.06	-0.14	-0.03	-0.03	0.98	0.98	0.99	0.99	-0.47	-0.45	-0.42	-0.40
Costa Rica	0.40	0.21	0.12	0.20	0.22	0.81	0.81	0.81	0.81	.	.	.	.
Croatia	0.28	-0.54	-0.58	-0.52	-0.51	0.64	0.64	0.64	0.64	.	.	.	.
Cyprus	0.66	0.34	0.14	0.27	0.29	0.75	0.79	0.80	0.80	.	.	.	.
Czech Republic	0.16	-0.20	-0.24	-0.21	-0.20	0.23	0.23	0.23	0.23	.	.	.	.
Côte d'Ivoire	0.14	-0.41	-0.44	-0.41	-0.41	0.88	0.87	0.87	0.87	0.12	0.13	0.14	0.15
Dem. Rep. Congo	0.27	-0.43	-0.46	-0.41	-0.40	0.95	0.95	0.95	0.95	-0.06	-0.06	-0.04	-0.03
Denmark	0.41	0.46	0.40	0.48	0.50	0.18	0.16	0.17	0.17	.	.	.	.
Djibouti	0.27	-0.23	-0.36	-0.31	-0.30	0.96	0.95	0.95	0.95	-0.29	-0.39	-0.38	-0.36
Dominican Rep.	0.68	0.22	0.13	0.27	0.31	0.88	0.88	0.89	0.89	.	.	.	.
Ecuador	0.41	-0.29	-0.35	-0.27	-0.26	0.85	0.85	0.85	0.85	.	.	.	.
Egypt	0.19	-0.30	-0.33	-0.29	-0.29	0.98	0.97	0.97	0.97	-0.12	-0.29	-0.28	-0.27
El Salvador	0.37	-0.26	-0.33	-0.25	-0.22	0.87	0.86	0.86	0.86	.	.	.	.
Estonia	0.43	-0.20	-0.37	-0.28	-0.27	0.18	0.18	0.18	0.18	.	.	.	.
Ethiopia	0.19	-0.33	-0.37	-0.33	-0.33	0.96	0.97	0.97	0.97	.	.	.	.
Finland	0.47	0.36	0.23	0.33	0.34	0.29	0.29	0.30	0.30	.	.	.	.
France	0.37	-0.03	-0.13	-0.06	-0.04	0.27	0.23	0.24	0.24	.	.	.	.
Gabon	0.35	-0.15	-0.21	-0.14	-0.14	0.89	0.89	0.89	0.89	0.28	0.28	0.30	0.30
Georgia	0.29	-0.39	-0.45	-0.39	-0.39	0.84	0.83	0.83	0.83	.	.	.	.
Germany	0.40	0.30	0.24	0.32	0.33	0.39	0.40	0.40	0.40	.	.	.	.
Ghana	0.29	0.18	0.11	0.16	0.18	0.95	0.94	0.95	0.95	0.31	0.32	0.34	0.35
Greece	0.28	-0.59	-0.62	-0.56	-0.56	0.71	0.72	0.73	0.73	.	.	.	.
Guatemala	0.42	0.30	0.16	0.24	0.27	0.90	0.90	0.91	0.91	.	.	.	.
Guinea	0.31	-0.07	-0.11	-0.05	-0.05	0.97	0.96	0.96	0.96	0.17	0.27	0.29	0.29
Haiti	0.39	0.16	-0.01	0.07	0.09	0.85	0.85	0.85	0.85	.	.	.	.
Honduras	0.45	0.11	-0.06	0.03	0.06	0.82	0.81	0.82	0.82	.	.	.	.
Hong Kong	0.51	0.45	0.29	0.40	0.43	0.25	0.24	0.24	0.24	.	.	.	.
Hungary	0.26	-0.24	-0.32	-0.27	-0.26	0.37	0.36	0.37	0.37	.	.	.	.
India	0.05	-0.20	-0.25	-0.24	-0.24	0.85	0.87	0.87	0.87	0.36	0.40	0.41	0.41
Indonesia	0.09	0.36	0.34	0.35	0.36	0.99	0.99	0.99	0.99	0.37	0.38	0.38	0.38
Iran	0.20	0.51	0.46	0.50	0.50	0.91	0.93	0.94	0.94	0.17	0.16	0.17	0.17
Iraq	0.22	-0.23	-0.23	-0.19	-0.19	0.83	0.83	0.84	0.84	-0.56	-0.69	-0.67	-0.67
Ireland	0.73	0.86	0.75	0.90	0.95	0.52	0.53	0.54	0.54	.	.	.	.
Israel	0.33	0.26	0.18	0.25	0.27	0.48	0.51	0.51	0.51	.	.	.	.
Italy	0.25	0.15	-0.01	0.04	0.05	0.69	0.69	0.69	0.69	.	.	.	.
Jamaica	0.68	0.24	0.11	0.25	0.25	0.84	0.80	0.81	0.81	.	.	.	.
Japan	0.14	-0.19	-0.22	-0.19	-0.19	0.26	0.26	0.26	0.26	.	.	.	.
Jordan	0.31	-0.36	-0.36	-0.30	-0.29	.	.	.	.	-0.22	-0.24	-0.22	-0.21
Kazakhstan	0.30	-0.20	-0.27	-0.21	-0.21	0.47	0.50	0.50	0.50	0.25	0.18	0.19	0.19
Kenya	0.23	0.20	0.16	0.21	0.22	0.94	0.94	0.94	0.94	0.24	0.22	0.24	0.24
Kosovo	0.40	-0.05	-0.12	-0.04	-0.03	0.89	0.89	0.90	0.90	0.39	0.43	0.45	0.47
Kuwait	0.36	0.26	0.35	0.42	0.42	0.94	0.91	0.92	0.92	-0.07	-0.08	-0.06	-0.06
Kyrgyzstan	0.37	-0.14	-0.20	-0.12	-0.13	0.69	0.69	0.70	0.70	0.07	-0.04	-0.01	-0.01

Laos	0.22	0.59	0.51	0.55	0.55	0.98	0.98	0.98	0.98	.	.	.	.
Latvia	0.46	-0.18	-0.27	-0.18	-0.16	0.37	0.39	0.39	0.39	.	.	.	.
Lebanon	0.45	0.00	-0.08	0.01	0.03	0.87	0.87	0.87	0.87	0.31	0.24	0.27	0.28
Lesotho	0.51	-0.04	-0.05	0.05	0.05	0.94	0.92	0.93	0.93	0.17	0.13	0.16	0.16
Liberia	0.35	0.52	0.46	0.53	0.55	0.93	0.92	0.93	0.93	0.40	0.41	0.43	0.44
Libya	0.28	0.46	0.44	0.50	0.50	.	.	.	.	.	.	.	.
Lithuania	0.39	-0.40	-0.46	-0.39	-0.37	0.41	0.40	0.41	0.41	.	.	.	.
Luxembourg	0.70	0.36	0.26	0.40	0.41	0.34	0.35	0.36	0.36	.	.	.	.
Macedonia	0.40	-0.32	-0.41	-0.33	-0.31	0.77	0.75	0.75	0.75	.	.	.	.
Madagascar	0.10	-0.40	-0.45	-0.43	-0.43	0.91	0.91	0.91	0.91	0.39	0.39	0.39	0.39
Malawi	0.27	0.20	0.14	0.20	0.20	0.98	0.98	0.99	0.99	0.10	0.07	0.09	0.10
Malaysia	0.13	-0.04	-0.07	-0.05	-0.04	0.93	0.93	0.93	0.93	0.08	0.05	0.06	0.06
Mali	0.27	-0.32	-0.36	-0.31	-0.31	0.96	0.95	0.96	0.96	-0.33	-0.39	-0.37	-0.37
Malta	0.58	0.45	0.31	0.43	0.47	0.87	0.86	0.87	0.87	.	.	.	.
Mauritania	0.38	-0.04	-0.13	-0.05	-0.05	0.99	0.98	0.99	0.99	-0.26	-0.26	-0.24	-0.22
Mauritius	0.47	0.40	0.32	0.41	0.41	0.95	0.94	0.95	0.95	0.55	0.54	0.57	0.57
Mexico	0.31	-0.03	-0.11	-0.04	-0.02	0.63	0.62	0.63	0.63	.	.	.	.
Moldova	0.54	-0.15	-0.24	-0.13	-0.14	0.74	0.72	0.73	0.73	.	.	.	.
Mongolia	0.29	0.18	0.11	0.17	0.17	0.45	0.47	0.47	0.47	.	.	.	.
Montenegro	0.29	-0.46	-0.54	-0.48	-0.47	0.60	0.59	0.59	0.59	.	.	.	.
Morocco	0.40	-0.37	-0.39	-0.31	-0.31	0.95	0.94	0.94	0.94	-0.23	-0.34	-0.31	-0.29
Mozambique	0.41	-0.31	-0.43	-0.35	-0.35	0.90	0.91	0.92	0.92	0.43	0.38	0.40	0.40
Myanmar	0.15	0.90	0.87	0.90	0.90	0.98	0.98	0.98	0.98	.	.	.	.
Nepal	0.39	-0.12	-0.19	-0.11	-0.11	0.95	0.95	0.95	0.95	.	.	.	.
Netherlands	0.41	0.73	0.65	0.73	0.75	0.32	0.33	0.34	0.34	.	.	.	.
New Zealand	0.75	0.83	0.71	0.86	0.92	0.32	0.28	0.29	0.29	.	.	.	.
Nicaragua	0.53	-0.13	-0.30	-0.20	-0.18	0.83	0.81	0.82	0.82	.	.	.	.
Niger	0.37	-0.33	-0.36	-0.28	-0.28	0.99	0.99	0.99	0.99	-0.64	-0.66	-0.64	-0.64
Nigeria	0.21	0.32	0.25	0.30	0.31	0.96	0.96	0.96	0.96	-0.02	-0.05	-0.04	-0.03
Oman	0.51	0.38	0.23	0.33	0.33	.	.	.	.	0.09	0.06	0.09	0.09
Pakistan	0.15	-0.07	-0.08	-0.05	-0.05	0.93	0.94	0.94	0.94	-0.32	-0.33	-0.33	-0.32
Palestina	0.32	-0.40	-0.47	-0.41	-0.40	0.94	0.93	0.94	0.94	-0.26	-0.28	-0.27	-0.26
Panama	0.28	0.09	-0.04	0.02	0.04	0.80	0.80	0.80	0.80	.	.	.	.
Paraguay	0.57	0.22	0.08	0.20	0.20	0.91	0.90	0.91	0.91	.	.	.	.
Peru	0.41	-0.15	-0.31	-0.23	-0.22	0.82	0.84	0.84	0.84	.	.	.	.
Philippines	0.50	0.37	0.24	0.34	0.36	0.95	0.94	0.95	0.95	.	.	.	.
Poland	0.32	-0.16	-0.23	-0.16	-0.15	0.67	0.69	0.69	0.69	.	.	.	.
Portugal	0.43	-0.22	-0.32	-0.24	-0.23	0.63	0.64	0.64	0.64	.	.	.	.
Qatar	0.68	0.45	0.40	0.53	0.53	0.96	0.97	0.98	0.98	0.13	0.12	0.16	0.16
Republic of Congo	0.39	-0.18	-0.23	-0.15	-0.15	0.89	0.87	0.87	0.87	-0.09	-0.07	-0.05	-0.05
Romania	0.40	-0.37	-0.49	-0.41	-0.40	0.80	0.81	0.81	0.81	.	.	.	.
Russia	0.17	-0.37	-0.41	-0.38	-0.38	0.33	0.32	0.32	0.32	.	.	.	.
Rwanda	0.13	-0.49	-0.53	-0.51	-0.51	0.96	0.96	0.96	0.96	0.48	0.49	0.50	0.50
Saudi Arabia	0.24	-0.05	-0.16	-0.11	-0.11	0.94	0.93	0.94	0.94	-0.12	-0.15	-0.14	-0.14
Senegal	0.37	-0.16	-0.22	-0.15	-0.14	0.97	0.97	0.97	0.97	0.01	0.02	0.04	0.06
Serbia	0.27	-0.50	-0.58	-0.52	-0.51	0.55	0.55	0.55	0.55	.	.	.	.
Sierra Leone	0.27	0.20	0.05	0.11	0.12	0.97	0.96	0.97	0.97	0.22	0.21	0.23	0.24
Singapore	0.34	-0.06	-0.12	-0.05	-0.05	0.59	0.56	0.57	0.57	.	.	.	.
Slovakia	0.30	-0.19	-0.23	-0.17	-0.16	0.50	0.50	0.50	0.50	.	.	.	.
Slovenia	0.45	0.24	0.18	0.27	0.29	0.44	0.46	0.46	0.46	.	.	.	.
Somalia	0.54	0.26	-0.02	0.09	0.12	0.99	0.99	0.99	0.99	-1.03	-1.19	-1.16	-1.14
South Africa	0.11	-0.09	-0.13	-0.11	-0.10	0.85	0.85	0.86	0.86	0.43	0.43	0.44	0.44
South Korea	0.26	0.07	-0.03	0.03	0.04	0.45	0.42	0.43	0.43	.	.	.	.
Spain	0.39	-0.06	-0.18	-0.11	-0.10	0.46	0.46	0.47	0.47	.	.	.	.
Sri Lanka	0.19	0.63	0.60	0.64	0.65	0.99	0.99	0.99	0.99	.	.	.	.
Sudan	0.33	0.17	0.09	0.16	0.17	0.95	0.94	0.94	0.94	-0.21	-0.33	-0.31	-0.30
Suriname	0.53	-0.01	-0.03	0.08	0.08	0.79	0.76	0.76	0.76	.	.	.	.
Swaziland	0.46	0.11	-0.03	0.07	0.07	0.94	0.93	0.93	0.93	0.14	0.10	0.13	0.13
Sweden	0.43	0.22	0.13	0.21	0.23	0.16	0.13	0.13	0.13	.	.	.	.
Switzerland	0.52	0.46	0.30	0.41	0.42	0.39	0.43	0.43	0.43	.	.	.	.
Syria	0.20	0.02	-0.03	0.01	0.01	0.84	0.84	0.84	0.84	-0.26	-0.33	-0.32	-0.31
Tajikistan	0.27	0.14	0.12	0.18	0.17	0.84	0.84	0.84	0.84	-0.26	-0.27	-0.26	-0.26

Tanzania	0.12	0.04	0.01	0.03	0.04	0.94	0.94	0.94	0.94	0.11	0.10	0.11	0.11
Thailand	0.10	0.41	0.40	0.42	0.42	0.98	0.98	0.98	0.98	.	.	.	.
Togo	0.24	-0.55	-0.58	-0.53	-0.53	0.86	0.87	0.87	0.87	0.27	0.25	0.26	0.26
Trinidad & Tobago	0.57	0.43	0.28	0.39	0.39	0.87	0.89	0.89	0.89	.	.	.	.
Tunisia	0.29	-0.23	-0.22	-0.16	-0.16	0.94	0.94	0.94	0.94	-0.09	-0.08	-0.06	-0.04
Turkey	0.17	-0.46	-0.49	-0.45	-0.45	0.83	0.83	0.83	0.83	0.41	0.39	0.40	0.41
Turkmenistan	0.20	0.51	0.46	0.50	0.50	0.80	0.80	0.81	0.81	0.32	0.27	0.28	0.28
Uganda	0.19	0.07	0.00	0.04	0.04	0.94	0.95	0.95	0.95	0.23	0.23	0.24	0.25
Ukraine	0.22	-0.31	-0.37	-0.32	-0.33	0.48	0.48	0.48	0.48	.	.	.	.
United Arab Emir.	0.32	0.15	0.09	0.16	0.16	0.93	0.93	0.93	0.93	0.08	0.12	0.13	0.14
United Kingdom	0.49	0.70	0.57	0.67	0.69	0.28	0.24	0.25	0.25	.	.	.	.
United States	0.22	0.86	0.77	0.82	0.82	0.66	0.66	0.67	0.67	.	.	.	.
Uruguay	0.39	-0.14	-0.27	-0.20	-0.19	0.38	0.37	0.38	0.38	.	.	.	.
Uzbekistan	0.14	0.23	0.23	0.26	0.26	0.59	0.59	0.59	0.59	0.21	0.22	0.23	0.23
Venezuela	0.17	-0.28	-0.33	-0.30	-0.29	0.79	0.79	0.79	0.79	.	.	.	.
Vietnam	0.07	-0.32	-0.33	-0.32	-0.32	0.32	0.31	0.31	0.31	.	.	.	.
Yemen	0.20	-0.37	-0.50	-0.46	-0.46	0.97	0.98	0.98	0.98	-0.30	-0.32	-0.31	-0.31
Zambia	0.17	0.12	0.05	0.09	0.09	0.96	0.95	0.96	0.96	0.36	0.39	0.40	0.40
Zimbabwe	0.46	-0.11	-0.19	-0.10	-0.09	0.88	0.87	0.88	0.88	0.30	0.20	0.22	0.24

Note: author calculation on Gallup World Poll Data. Column (1) shows the average share of people with connection abroad in the country. Columns (2), (6) and (10) show the average level of each cultural trait. Columns (3), (7) and (11) show the average level of each cultural trait in the population without connection. Columns (4), (8) and (12) show the predicted level of each cultural trait in the population after including the effect of the connection estimated in Table 6.3. Columns (5), (9) and (13) shows the predicted level of each cultural trait in the population after including the effect of the connection and interaction with the cultural trait in the destination country estimated in Table 6.3

Figure A9: Average Trait by country of birth - Population and Migrants abroad



Note: authors' calculations on Gallup World Poll Data. The figure plots the average cultural trait for the population in the origin countries and their migrants abroad.

General Conclusions

This PhD thesis aims to increase the understanding of the role of international migration on individuals and societies, through its effects on population diversity, on natives' voting preferences and on individual cultural traits and beliefs.

Chapter 1 and 2 show, both across the US states and across 100 countries, that immigration-driven diversity has positive economic implications. By increasing the skill and knowledge set of a country, diversity can enhance economic growth and countries economic complexity when potential complementarities between skills are available. In particular, diversity positively influence R&D sector and countries diversification in their export basket.

Investigating the implication of immigration on European voters' preferences on 12 countries over 2007-2016, Chapter 3 and 4 show that the effect of immigration on natives voting preferences is education-specific, both for migrants and voters. The presence of highly educated immigrants, who are easily integrated into the labour market and perceived as contributors to the welfare system, has pushed European voters towards less nationalistic and more redistributive parties. Lowly educated immigration has instead driven voters towards more nationalistic and less redistributive parties. Moreover, we show that the presence of immigrants only influences lowly educated natives voting behaviour, while has almost no effect on college educated natives.

Finally, Chapter 5 and 6 shed new light on the relation between culture and international migration. Focusing on individuals in the MENA region, I show that emigrants are culturally selected: less religious and more gender-egalitarian individuals are more prone to move towards OECD destination countries. Then, I show that having a reliable connection abroad makes natives culturally different. Sharing the knowledge and experience of people around the world makes individuals more active in the society, gender-egalitarian and religious. By moving from their origin countries and interacting with their peers at home, migrants influence cultural traits both in the origin and destination countries.

Bibliography

Aakvik, A. (2001). Bounding a matching estimator: the case of a Norwegian training program. *Oxford Bulletin of Economics and Statistics* 63(1): 115-143.

Adsera, A. & M. Pytlikova (2015). The role of Language in shaping international migration. *Economic Journal* 125(586): F49-F81

Ager, P., and Brückner, M. (2013). Cultural diversity and economic growth: Evidence from the US during the age of mass migration. *European Economic Review*, 64:76–97.

Ager, P., & Ciccone, A. (2017). Agricultural risk and the spread of religious communities. *Journal of the European Economic Association* 16(4): 1021-1068.

Alesina A., R. Baqir & W. Easterly (1999). Public Goods and Ethnic Divisions *The Quarterly Journal of Economics* 114(4): 1243–1284.

Alesina, A., E. Glaeser & B. Sacerdote (2001). Why Doesn't the United States Have a European-Style Welfare State? *Brookings Paper on Economics Activity*: 187-278.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, 8(2): 155-194.

Alesina, A., and La Ferrara, E. (2005). Ethnic diversity and economic performance. *Journal of Economic Literature*, 43(3):762–800.

Alesina, A. & P. Giuliano (2010). The Power of the Family. *Journal of Economic Growth* 15(2): 93–125

Alesina, A., & Giuliano, P. (2015). Culture and institutions. *Journal of Economic Literature* 53(4): 898-944.

Alesina, A., Harnoss, J. and Rapoport, H. (2016). Birthplace diversity and economic prosperity. *Journal of Economic Growth*, 21:101–138.

Alesina, A. & A. Miano & S. Stantcheva (2018). Immigration and Redistribution. *NBER Working paper* 24733.

Alesina, A., J. Harnoss & H. Rapoport (2018b). Immigration and the Future of the Welfare State in Europe *PSE Working Papers* 2018-04.

Alesina, A., Murard, E., and Rapoport, H. (2019). Immigration and Preferences for Redistribution in Europe National Bureau of Economic Research. WP No. w25562

Alexander, J.C. (2008). Clifford Geertz and the Strong Program: The Human Sciences and Cultural Sociology. *Cultural Sociology* 2: 157

Alexander, A. & C. Welzel (2011). Islam and Patriarchy: How Robust is Muslim Support for Patriarchal Values? *International Review of Sociology* 21(2): 249–276

Algan, Y. & P. Cahuc (2010). On the Origin of Gender Roles: Women and Plogh. *Quarterly Journal of Economics* 128(2): 469-530

Algan, Y. & S. Guriev & E. Papaioannou & E. Passari (2017). The European trust crisis and the rise of populism. *CEPR Discussion Paper* DP12444.

Altonji, J. G. & Card, D. (1991). The effects of immigration on the labor market outcomes of less-skilled natives. In *Immigration, trade, and the labor market* (pp. 201-234). University of Chicago Press.

Altonji, J.G. & T.E. Elder & C.R. Taber (2005). Selection on Observed and Unobserved Variables: assessing the effectiveness of Catholic Schools. *Journal of Political Economy* 113(1): 151-184

Altonji, J. G., Conley, T., Elder, T. E., & Taber, C. R. (2010). Methods for using selection on observed variables to address selection on unobserved variables. Mimeo, Yale University.

Anelli, M., & Peri, G. (2017). Does emigration delay political change? Evidence from Italy during the Great Recession. *Economic Policy* 32(91): 551-596.

Angrist J, and Pischke, J-S. (2009). Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press (392 p.).

Angrist, J. D., & Pischke, J. S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives* 24(2): 3-30.

Asselin, L. (2002). Multidimensional poverty theory. Composite indicator of multidimensional poverty. Levis: Institut de Mathematique Gauss.

Artuc, E., Docquier, F., Özden, C., and Parsons, C. (2015). A Global Assessment of Human Capital Mobility: The Role of Non-OECD Destinations. *World Development*, 65(1):6–26.

Arzheimer, K. (2009). Contextual Factors and the Extreme Right Vote in Western Europe, 1980-2002 *American Journal of Political Science* 53: 259-275.

Ashraf, Q., and Galor, O. (2013). The Out of Africa hypothesis, Human genetic diversity, and comparative economic development. *American Economic Review*, 103(1):1–46.

Ashraf, Q., Galor, O., and Klemp, M. (2015). The Out of Africa hypothesis of comparative development reflected by nighttime light intensity. Working Paper No. 2014-4, Brown University, Departments of Economics.

Atkin, D. (2016). The caloric costs of culture: Evidence from Indian migrants. *American Economic Review* 106(4): 1144-81.

Atkin, D., Colson-Sihra, E., & Shayo, M. (2019). How Do We Choose Our Identity? A Revealed Preference Approach Using Food Consumption. *NBER Working Paper* No. 25693.

Aubry, A., Burzyński, M., and Docquier, F. (2016). The welfare impact of global migration in OECD countries. *Journal of International Economics* , 101:1–21.

Baas, T., & Brücker, H. (2012). The macroeconomic consequences of migration diversion: Evidence for Germany and the UK. *Structural Change and Economic Dynamics* 23(2): 180-194.

Bahar, D., Hauptmann, A., Özgüzel, C., and Rapoport, H. (2018). Let their Knowledge Flow: The Effect of Returning Refugees on Export Performance in the Former Yugoslavia. CESifo Working Paper. No. 7371

Bahar, D., and Rapoport, H. (2018). Migration, knowledge diffusion and the comparative advantage of nations. *The Economic Journal*, 128(612):F273-F305.

Bahar, D., and Santos, M. A. (2018). One more resource curse: Dutch disease and export concentration. *Journal of Development Economics*, 132:102-114.

Bahar, D., Choudhury, P., & Rapoport, H. (2019). Migrant Inventors and the Technological Advantage of Nations. HBS Working Paper 19-119

Bailey, M., Cao, R., Kuchler, T., Stroebel, J., & Wong, A. (2018). Social connectedness: measurement, determinants, and effects. *Journal of Economic Perspectives* 32(3): 259-80.

Balassa, B. (1965). Trade liberalisation and "revealed" comparative advantage. *The Manchester School*, 33(2): 99-123.

Barone, G. & F. D'Acunto & G. Narciso (2015). Telecracy: Testing for channels of persuasion. *American Economic Journal: Economic Policy* 7(2): 30–60.

Barone, G. & A. D'Ignazio & G. De Blasio & P. Naticchioni (2016). Mr. Rossi, Mr. Hu and politics. The role of immigration in shaping natives' voting behavior. *Journal of Public Economics* 136: 1–13.

Barro, R & J. Lee (2013). A New Data Set of Educational Attainment in the World, 1950-2010 *Journal of Development Economics*, 104:184-198

Barsbai T., Rapoport, H., Steinmayr, A., and Trebesch, C. (2017). The effect of labor migration on the diffusion of democracy: Evidence from a former Soviet Republic. *American Economic Journal: Applied Economics*, 9(3):36–69.

Batista, C., & Vicente, P. C. (2011). Do migrants improve governance at home? Evidence from a voting experiment. *The World Bank Economic Review* 25(1), 77-104.

Batista, C., Seither, J., & Vicente, P. C. (2019). Do migrant social networks shape political attitudes and behavior at home? *World Development* 117: 328-343.

Battu, H. & Y. Zenou (2010). Oppositional Identities and Employment for Ethnic Minorities: Evidence from England. *Economic Journal* 120(542): F52–F71

Baudin, T. (2010). A role for cultural transmission in fertility transitions. *Macroeconomic Dynamics* 14(4): 454-481.

Baxter, J. & E.W. Kane (1995). Dependence and independence: A cross-national analysis of gender inequality and gender attitudes. *Gender & Society* 9(2): 193–215

Becerra, D. & M. Gurrola & C. Ayon & D. Androff & J. Krysik & K. Gerdes & L. Moya-Salas & E. Segal (2010). Poverty and Other Factors Affecting Migration Intentions among Adolescents in Mexico. *Journal of Poverty* 14: 1-16

Becerra, D. (2012). The Impact of anti-immigration policies and perceived discrimination in the United States on migration intentions among Mexican adolescents. *International Migration* 50(4): 20-32

Becker, S. O., & Caliendo, M. (2007). Sensitivity analysis for average treatment effects. *The Stata Journal* 7(1): 71-83.

Becker, S.O. & T. Fetzer (2017). Does Migration Cause Extreme Voting? *Center for Competitive Advantage in the Global Economy and The Economic & Social Research Council.*

Beine, M., Docquier, F., and Schiff, M. (2013). International migration, transfer of norms and home country fertility. *Canadian Journal of Economics*, 46(4):1406–1430.

Belot, M. & S. Ederveen (2012). Cultural barriers in migration between OECD countries. *Journal of Population Economics* 25(3): 1077–1105

Belot, M & T.J. Hatton (2012). Immigrant Selection in the OECD. *Scandinavian Journal of Economics* 114(4): 1105–1128

Benabou, R. & D. Ticchi & A. Vindigni (2015). Religion and Innovation. *American Economic Review* 105(5) : 346-351

Bergh, J. (2007). Gender Attitudes and Modernization Processes. *International Journal of Public Opinion Research* 19(1): 5–23

Berlinschi, R., & Harutyunyan, A. (2018). Do migrants think differently? Evidence from Eastern European and post-Soviet states. *International Migration Review* forthcoming.

Bertoli, S. and Marchetta, F. (2015). Bringing it all back home–return migration and fertility choices. *World Development*, 65:27–40.

Bertoli, S. & I. Ruyssen (2018). Networks and Migrants Intended Destination. *Journal of Economic Geography* 18(4): 705-728

Bertoli, S., Docquier, F., Rapoport, H., & Ruyssen, I. (2019). Weather shocks and migration intentions in Western Africa: Insights from a multilevel analysis. Mimeo

Bisin, A. & T. Verdier (1998). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory* 97: 298-319

Bisin, A., & Verdier, T. (2000). "Beyond the melting pot": cultural transmission, marriage, and the evolution of ethnic and religious traits. *The Quarterly Journal of Economics* 115(3): 955-988.

Bisin, A. & E. Pataccini & T. Verdier & Y. Zenou (2011). Ethnic Identity and Labour Market Outcomes of Immigrants in Europe. *Economic Policy* 25(65): 57–92

Boeri, T. (2002). Let Social Policy Models Compete and Europe Will Win, conference in the John F. Kennedy School of Government, Harvard University, 11–12 April 2002.

Boeheim, R., Horvath, G., and Mayr, K. (2012). Birthplace diversity of the workforce and productivity spill-overs in firms. WIFO Working PapersNo. 438.

Borjas, G.J. (1994). The Economics of Immigration. *American Economic Review*, 32:1667–1717.

Borjas, G.J. (2003). The labour demand curve is downward sloping: reexamining the impact of immigration on the labour market. *Quarterly Journal of Economics*, 118(4):1335–1374.

Borjas, G.J. (2015). Immigration and Globalization: A review Essay. *Journal of Economic Literature*, 53(4):961–974.

Borjas, G.J. (2017). The labor supply of undocumented immigrants. *Labour Economics*, 46: 1-13.

Borjas, G. J. (2019). Immigration and Economic Growth. National Bureau of Economic Research. WP No. w25836.

Bouoiyour, J & A. Miftah (2017). Do Migrants Transfer Political and Cultural Norms to Their Origin Country? Some Evidence from Some Arab Countries Economic Research Forum (ERF) Working Paper No: 1098.

Boughzala, M. & S. Romdhane (2017). Tunisia: The Prospects for Democratic Consolidation. in: Ibrahim Elbadawi and Samir Makdisi (eds.) Democratic Transitions in the Arab World, Cambridge, UK: Cambridge University Press.

Bove, V., & Elia, L. (2016). Migration, diversity, and economic growth. *World Development*, 89:227-239.

Braga, M., Checchi, D., and Meschi, E. (2013). Educational policies in a long-run perspective. *Economic Policy*, 28(73): 45-100.

Brunner, B. & A. Kuhn (2018). Immigration, Cultural Distance and Natives' Attitudes Towards Immigrants: Evidence from Swiss Voting Results *Kyklos* 71(1): 25-58.

Budge, I., Klingemann, H. D., Volkens, A., Bara, J., Tanenbaum, E., Fording, R. C., Hearl, D. J, Kim, H. M., McDonald M. & Silvia M. (2001). Mapping Policy Preferences. Estimates for Parties, Electors, and Governments 1945-1998. Oxford University Press.

Burzyński, M., Docquier, F., and Rapoport, H. (2018). The changing structure of immigration to the OECD: what welfare effects on member countries? *IMF Economic Review*, 66(3): 564-601.

Cadot, O., Carrère, C., and Strauss-Kahn, V. (2011). Export diversification: What's behind the hump? *Review of Economics and Statistics*, 93(2):590-605.

Cadot, O., Carrere, C., & Strauss-Kahn, V. (2013). Trade diversification, income, and growth: what do we know?. *Journal of Economic Surveys*, 27(4):790-812.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys* 22(1): 31-72.

Campante, F. & D. Yanagizawa-Drott (2015). Does Religion Affect Economic Growth and Happiness? Evidence from Ramadan. *Quarterly Journal of Economics* 130(2): 615-658

Card, D. (2001). Immigration Inflows, Native Outflows and the Local Labor Market Impacts of Higher Immigration. *Journal of Labor Economics* 19: 22-64.

Card, D. & C. Dustmann & I. Preston (2006). Understanding attitudes to immigration: The migration and minority module of the first European Social Survey. In: R. Jowell & R. Fitzgerald & G. Eva (Eds.), Measuring attitudes cross-nationally – lessons from the European Social Survey. Sage Publication.

Card, D. (2009). Immigration and inequality. *American Economic Review*, 99(2):1–21.

Card, D (2012). Comment: The Elusive search for Negative Wage Impacts of Immigration *Journal of the European Economic Association*, 10(1):211-215.

Card, D. & C. Dustmann & I. Preston (2012). Immigration, Wages and Compositional Amenities. *Journal of the European Economic Association* 10(1): 78–119.

Card, D. & Peri, G. (2016). Immigration economics by George J. Borjas: a review essay. *Journal of Economic Literature*, 54(4): 1333-49.

Casey, T. & C. Dustmann (2010). Immigrants' Identity, Economic Outcomes and the Transmission of Identity Across Generations. *Economic Journal* 120(2): F31–F51

Chabé-Ferret, B. (2019). Adherence to cultural norms and economic incentives: evidence from fertility timing decisions. *Journal of Economic Behavior & Organization* 162: 24-48.

Chase, G. (2014). Religiosity, corruption and economic growth. *Quarterly Review of Business Disciplines* 1(3): 263-272

Chauvet, L., & Mercier, M. (2014). Do return migrants transfer political norms to their origin country? Evidence from Mali. *Journal of Comparative Economics* 42(3): 630-651.

Chevalier, A., Elsner, B., Lichter, A. & Pestel, N. (2018). Immigrant Voters, Taxation and the Size of the Welfare State. *IZA DP* No. 11725.

Chort I. (2014). Mexican Migrants to the US: What do Unrealized Migration Intentions Tell Us about Gender Inequalities? *World Development* 59: 535-552

Clemens, M.A. & L. Pritchett (2016). The new economic case for migration restrictions: an assessment. Manuscript (Center for Global Develomment).

Coleman, J. S. (1990). Foundations of social theory. Harvard university press.

Collier, P. (2013). Exodus: How migration is changing our World. Oxford University Press (309 p.).

Collier P. (2017). Culture, Politics, and Economic Development. *Annual Review of Political Science* 20: 111-125

Constantin, A., & Voicu, M. (2015). Attitudes towards gender roles in cross-cultural surveys: Content validity and cross-cultural measurement invariance. *Social Indicators Research* 123(3): 733-751.

Costello, A.B. & J.W. Osborne (2005). Best Practices in Exploratory Factor Analysis: Four recommendations for getting the most from your Analysis. *Practical Assessment, Research & Evaluation* 10(7): 1-9

Dahlberg, M. & K. Edmark & H. Lundqvist (2012). Ethnic Diversity and Preferences towards redistribution. *Journal of Political Economy* 120(1): 41–76.

Dao, T.H. & F. Docquier & C. Parsons & G. Peri (2018). Migration and Development: Dissecting the Anatomy of the Mobility Transition *Journal of Development Economics* 132: 88-101

De Jong, G.F. & K. Richter & P. Isarabhakdi (1996). Gender, Values and Intention to Move in Rural Thailand. *International Migration Review* 30(3): 748-770

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics* 84(1): 151-161.

Della Vigna, S. & M. Gentzkow (2010). Persuasion: Empirical Evidence *Annu. Rev. Econ.*, 2(1): 643-669.

Desmet, K. & M. Le Breton & I. Ortuño-Ortín & S. Weber (2011). The stability and breakup of nations: a quantitative analysis. *Journal of Economic Growth* 16(3): 183-213

Desmet, K., Ortuño-Ortín, I., & Wacziarg, R. (2012). The political economy of linguistic cleavages. *Journal of development Economics*, 97(2):322-338.

Desmet, K., Ortuño-Ortín, I., & Wacziarg, R. (2017). Culture, ethnicity, and diversity. *American Economic Review* 107(9): 2479-2513.

Desmet, K., & Wacziarg, R. (2018). The cultural divide. *NBER Working Paper* No. 24630.

di Giovanni, J. & A.A. Levchenko & F. Ortega (2015). A global view of cross-border migration. *Journal of the European Economic Association*, 13(1):168-202.

DiPrete, T. A., & Gangl, M. (2004). Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments. *Sociological Methodology* 34(1): 271-310.

Dibeh, G. & A. Fakih & W. Marrouch (2018). Decision to Emigrate amongst the Youth in Lebanon. *International Migration* 56(1): 5-22

Di Tella, R. & R. MacCulloch (2014). Culture, Beliefs and Economic Performance. Motu Working Paper No. 14-06

Diwan, I., Tuccio, M., & J. Wahba (2017). Diaspora Networks as a bridge between civilizations. *World Scientific Encyclopaedia of Global Migration* forthcoming.

Diwan, I. & Z. Tzannatos & T. Akin (2018). Debunking the Myth: Economic Values in the Arab World Through the Prism of Opinion Polls. *Middle East Development Journal* 10(1): 31-63

Docquier, F. & O. Lohest & A. Marfouk (2007). Brain drain in developing countries. *World Bank Economic Review* 21: 193–218.

Docquier, F. & C. Ozden & G. Peri (2014). The Labour Market Effects of Immigration and Emigration in OECD countries. *The Economic Journal* 124: 1106-1145.

Docquier, F., G. Peri, I. Ruyssen (2014). The Cross-country Determinants of Potential and Actual Migration. *International Migration Review* 48: S37–S99.

Docquier, F., Machado, J., and Sekkat, K. (2015). Efficiency gains from liberalizing labor mobility. *The Scandinavian Journal of Economics*, 117(2):303-346.

Docquier, F., Lodigiani, E., Rapoport, H., and Schiff, M. (2016). Emigration and democracy. *Journal of Development Economics*, 120(1):209–223.

Docquier, F., Turati, R., Valette, J., & Vasilakis, C. (2019b). Birthplace Diversity and Economic Growth: Evidence from the US States in the Post-World War II Period. *Journal of Economic Geography* forthcoming.

Docquier, F., Tansel, A., & Turati, R. (2019). Do Emigrants Self-Select Along Cultural Traits? Evidence from the MENA Countries. *International Migration Review* forthcoming.

Dorn, D. (2009). Essays on Inequality, Spatial Interaction, and the Demand for Skills. Dissertation University of St. Gallen no. 3613, September 2009.

Doz, C. & D. Giannone & L. Reichlin (2012). A Quasi-maximum likelihood approach for large, approximate dynamic factor models. *Review of Economics and Statistics* 94(4): 1014-1024

Drinkwater, S. & Ingram, P. (2009). How different are the British in their willingness to move? Evidence from international social survey data. *Regional Studies* 43(2): 287-303

Duflo, E. (2012). Women's Empowerment and Economic Development. *Journal of Economic Literature* 50(4): 1051-79

Dustmann, C. & A. Okatenko (2014). Out-migration, wealth constraints, and the quality of local amenities. *Journal of Development Economics* 110: 52-63

Dustmann, C. & T. Frattini (2014). The fiscal effects of immigration to the UK. *The Economic Journal*, 124(580):F593-F643.

Dustmann, C., Vasiljeva, K., & Damm, A. P. (2018). Refugee migration and electoral outcomes. *Review of Economic Studies* forthcoming.

Edo, A. & Giesing, Y. & J. Öztunc & P. Poutvaara (2019). Immigration and Electoral Support for the Far Left and Far Right. *European Economic Review* 115: 99-143

Eger, M. A. (2009). Even in Sweden: the effect of immigration on support for welfare state spending. *European Sociological Review*, 26(2): 203-217.

Elbadawi, I. & S. Makdisi (2017). Introduction and Summary. in: Ibrahim Elbadawi and Samir Makdisi (eds.) Democratic Transitions in the Arab World, Cambridge, UK: Cambridge University Press.

El Mikawy, N. & M. Mohieddin & S. El Ashmaouy (2017). Egypt: The Protracted Transition from Authoritarianism to Democracy and Social Justice. in: Ibrahim Elbadawi and Samir Makdisi (eds.) Democratic Transitions in the Arab World, Cambridge, UK: Cambridge University Press.

Facchini, G. & A.M. Mayda (2009). Does the welfare state affect individual attitudes towards immigrants? Evidence across countries *The Review of Economics and Statistics* 91(2): 295-314.

Falco, C. & V. Rotondi (2016). The less extreme, the more you leave: radical Islam and Willingness to migrate. *World Development* 88: 122-133

Falck, O. & S. Heblich & A. Lameli & J. Suedekum (2012). Dialects, Cultural Identity and Economic Exchange. *Journal of Urban Economics* 72: 225-239

Falck, O. & A. Lameli & J. Ruhouse (2018). Cultural Biases in Migration: Estimating Nonmonteary Migration Costs. *Papers in Regional Science* 97(2): 411-438

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics* 133(4): 1645-1692.

Felbermayr, G. & F. Toubal (2010). Cultural Proximity and Trade. *European Economic Review* 54(2): 279-291

Felipe, J., Kumar, U., Abdon, A., & Bacate, M. (2012). Product complexity and economic development. *Structural Change and Economic Dynamics*, 23(1):36-68.

Fernandez, R. & A. Fogli (2009). Culture: an empirical Investigation of Beliefs, Work and Fertility. *American Economic Journal: Macroeconomics* 1(1): 146-177

Feenstra, R. C., Lipsey, R. E., Deng, H., Ma, A. C., & Mo, H. (2005). World trade flows: 1962-2000. National Bureau of Economic Research WP No. w11040.

Feenstra, R. C., Inklaar, R. & Timmer, M. P. (2015). The Next Generation of the Penn World Table. *American Economic Review* 105(10): 3150-3182.

Feyrer, J. (2019). Trade and income: Exploiting time series in geography. *American Economic Journal: Applied Economics*, Forthcoming.

Flanagan, C., & Levine, P. (2010). Civic engagement and the transition to adulthood. *The future of children* 159-179.

Fukuyama, F. (2001). Social capital, civil society and development. *Third World Quarterly* 22(1): 7-20.

Fulford, S., Petkov, I., and Schiantarelli, F. (2017). Does it matter where you came from? Ancestry composition and economic performance of U.S. counties, 1850-2010. Boston College Working Papers in Economics No. 875.

Gallup (2010). Measuring the State of Muslim-West Relations – Assessing the "New Beginning." Abu Dhabi Gallup Center.

Galor, O., & Savitskiy, V. (2018). Climatic roots of loss aversion. *NBER Working Paper* No. 25273.

Giavazzi, F., Petkov, I., & Schiantarelli, F. (2019). Culture: Persistence and evolution. *Journal of Economic Growth* 24(2): 117-154.

Gill, A. (1999). Government Regulation, Social Anomie and Protestant Growth in Latin America: a cross-national analysis. *Rationality and Society* 11: 287-316

Giuliano, P., & Nunn, N. (2019). Understanding Cultural Persistence and Change. *NBER Working Paper* No. 23617.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2018). Bartik Instruments: What, When, Why, and How. National Bureau of Economic Research. WP No. w24408

Gorodnichenko, Y. & G. Roland (2017). Culture, Institutions and the Wealth of Nations. *The Review of Economics and Statistics* 99(3): 402-416

Granovetter, M. (2005). The impact of social structure on economic outcomes. *Journal of Economic Perspectives* 19(1): 33-50.

Greenberg, J. (1956) The measurement of linguistic diversity. *Language*, 32:109-115

Greif, A., & Tabellini, G. (2010). Cultural and institutional bifurcation: China and Europe compared. *American Economic Review* 100(2): 135-40.

Grogger, J. & G.H. Hanson (2011). Income maximization and the selection and sorting of international migrants. *Journal of Development Economics* 95(1): 42–57

Guiso, L., Sapienza, P., & Zingales, L. (2006). Does culture affect economic outcomes? *Journal of Economic Perspectives* 20(2): 23-48.

Guiso, L. & P. Sapienza & L. Zingales (2009). Cultural Biases in Economic Exchange? *Quarterly Journal of Economics* 124(3): 1095–1131

Hainmueller, J. & M.J. Hiscox (2007). Educated preferences: Explaining attitudes toward immigration in Europe. *International Organization* 61(2): 399-442.

Hainmueller, J. & M.J. Hiscox (2010). Attitudes toward highly skilled and low-skilled immigration: Evidence from a survey experiment. *American Political Science Review* 104(1): 61-84.

Halla, M. & A. F. Wagner & J. Zweimuller (2017). Immigration and voting for far right. *Journal of the European Economic Association* 15(6): 1341–1385.

Hammel, E.A. (1990). A Theory of Culture for Demography. *Population and Development Review* 16(3): 455-485

Hanson, G.H. & K. Scheve & M.J. Slaughter (2007). Public Finance and individual preferences over globalization strategies. *Economics & Politics* 19(1): 1-33.

Hangartner, D., Dinas, E., Marbach, M., Matakos, K., & Xefteris, D. (2019). Does exposure to the refugee crisis make natives more hostile? *American Political Science Review* 113(2): 442-455.

Harmon, N.A. (2018). Immigration, Ethnic Diversity and Political Outcomes: Evidence from Denmark. *Journal of Scandinavian Economics*, 120(4): 1043-1074.

Hatton, T. (2016). Immigration, public opinion and the recession in Europe. *Economic Policy* 31(86): 205-246.

Hausmann, R., and Klinger, B. (2007). The structure of the product space and the evolution of comparative advantage. Center for International Development at Harvard University, WP No. 146

Hausmann, R., and Hidalgo, C. A. (2011). The network structure of economic output. *Journal of Economic Growth*, 16(4):309-342.

Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., Simoes, A., and Yildirim, M. A. (2014). The atlas of economic complexity: Mapping paths to prosperity. Mit Press.

Head, K., Mayer, T. and Ries, J. (2010) The erosion of colonial trade linkages after independence. *Journal of International Economics*, 81(1):1-14

Heckman, J. , Ichimura, H., & Todd, P. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies* 64(4): 605-654.

Heckman, J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing Selection Bias Using Experimental Data. *Econometrica* 66(5): 1017-1098.

Hellevik, O. (2009). Linear versus logistic regression when the dependent variable is a dichotomy. *Quality & Quantity* 43(1): 59-74.

Hidalgo, C. A., Klinger, B., Barabási, A. L., and Hausmann, R. (2007). The product space conditions the development of nations. *Science*, 317(5837):482-487.

Hidalgo, C. A., and Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the national academy of sciences*, 106(26):10570-10575.

Hjort, J. (2014). Ethnic divisions and production in firms. *Quarterly Journal of Economics*, 129(4):1899-1946.

Hix, S. & Lord, C. (1997). Political parties in the European Union. Macmillan International Higher Education.

Hoffman, S. & F. Marsiglia & S. Ayers (2015). Religiosity and migration aspirations among Mexican youth. *Journal of International Migration and Integration* 16: 173-186

Hunt, J. and M. Gauthier-Loisells (2010). How much does immigration boost innovation. *American Economic Journal: Macroeconomics*, 2:31-56

Ichino, A., Mealli, F., & Nannicini, T. (2008). From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity? *Journal of Applied Econometrics* 23(3): 305-327.

Imbens, G. & D. Rubin (2012). Casual Inference for Statistic, Social and Biomedical Sciences. Cambridge University Press

Imbs, J., and Wacziarg, R. (2003). Stages of diversification. *American Economic Review*, 93(1):63-86.

Inglehart, R., & Baker, W. E. (2000). Modernization, cultural change, and the persistence of traditional values. *American Sociological Review* 65: 19-51.

Inglehart, R. F., Ponarin, E., & Inglehart, R. C. (2017). Cultural change, slow and fast: The distinctive trajectory of norms governing gender equality and sexual orientation. *Social Forces* 95(4): 1313-1340.

Inglehart, R. (2018). Cultural evolution: people's motivations are changing, and reshaping the world. Cambridge University Press.

Iranzo, S. and G. Peri (2009). Migration and Trade: Theory with an application to the eastern-western European integration. *Journal of International Economics*, 79(1):1-19

Islam, A.& P.A. Raschky (2013). Cultural Distance, Immigrants' Identity, and Labour Market Outcomes. Discussion Paper 21/13. Monash University.

Jackson, M. O. (2014). Networks in the understanding of economic behaviors. *Journal of Economic Perspectives* 28(4): 3-22.

Jaeger, D. A., Ruist, J., and Stuhler, J. (2018). Shift-share instruments and the impact of immigration. National Bureau of Economic Research. WP No. w24285.

Javorcik, B. S., Lo Turco, A., & Maggioni, D. (2018). New and improved: Does FDI boost production complexity in host countries?. *The Economic Journal*, 128(614):2507-2537.

Jónsson G. (2008). Migration aspirations and immobility in a Malian Soninke village. IMI Working Paper 10-2008, International Migration Institute, UK, Oxford.

Just, P. & J. Monaghan (2000). Social and Cultural Anthropology: A Very Short Introduction. Oxford, UK: Oxford University Press.

Kahane, L., Longley, N., and Simmons, R. (2013). The effects of coworker heterogeneity on firm-level output: Assessing the impacts of cultural and language diversity in the national hockey league. *The Review of Economics and Statistics*, 95(1):302–314.

Kaestner, R. & O. Malamud (2014). Self-selection and International Migration: New Evidence from Mexico. *Review of Economics and Statistics* 96(1): 78-91

Kammoun, R. (2014). NGOs and Women's Rights in Tunisia. in: Galia Golan and Walid Salem (eds.) Non-State Actors in the Middle East: Factors for Peace and Democracy, University of California at Los Angeles Center for Middle East Development. New York, ny, USA: Rutledge.

Kashavarzian, A. (2010). Iran In Angrist (ed.) Politics and Society in the Contemporary Middle East, Boulder, Colorado, USA: Lynne Rienner Publishers.

Karouby, C. (2014). The Role of the Maronite Diaspora. In: G. Golan and W. Salem (eds.) Non-State Actors in the Middle East: Factors for Peace and Democracy, London, UK: Routledge, Taylor and Francis Group.

Kerr, W. R. & W. F. Lincoln (2010). The supply side of innovation: H-1B visa reforms and US ethnic invention. *Journal of Labour Economics* 28(3): 473-508.

Klingemann, H. D., Volkens, A., Bara, J., Budge, I. & M. McDonald (2006). Mapping Policy Preferences II: Estimates for Parties, Electors and Governments in Eastern Europe, European Union and OECD 1990–2003. Oxford University Press

Knack, S. & P. Keefer (1997). Does Social Capital Have an Economic Payoff? A Cross-Country Investigation. *Quarterly Journal of Economics* 112(4): 1251-1288

Kostenko, V. & E. Ponarin & M. Shteiwi & O. Strebkova (2017). Historical Legacies and Gender Attitudes in the Middle East. Economic Research Forum (ERF) Working Paper No: 1105

Krieger, T. & L. Renner & J. Ruhose (2018). Long-term relatedness between Countries and International Migrant Selection. *Journal of International Economics* 113: 35-54

Kuran, T. (2012). The Long Divergence: How Islamic Law Held Back the Middle East. Princeton University Press.

Lim G.C. & V.H. Nguyen (2015). Alternative weighting approaches to computing indexes of economic activity. *Journal of Economic Surveys* 29(2): 287-300

Levitt, P. (1998). Social remittances: Migration driven local-level forms of cultural diffusion. *International Migration Review* 32(4): 926-948.

Lodigiani, E. and Salomone, S. (2012). Migration-induced transfers of norms: The case of female political empowerment. *IRES Discussion Papers 2012–1.*

Lugo, L., & Cooperman, A. (2011). The future of the global Muslim population. Pew Research Center, USA.

Lyons, E. (2017). Team production in international labor markets: Experimental evidence from the field. *American Economic Journal: Applied Economics*, 9(3): 70-104.

Manacorda, M., A. Manning and J. Wadsworth (2012). The impact of immigration on the structure of wages: theory and evidence from Britain. *Journal of the European Economic Association*, 10(1):120-151

Manchin, M. & R. Manchin & S. Orazbayev (2014). Desire to Migrate Internationally and Locally and the Importance of Satisfaction with Amenities. Paper presented at the FIW-wiiw Seminars in International Economics, 10 April 2014, Austria, Vienna.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the national cancer institute* 22(4): 719-748.

Marshall, M. G., Gurr, T. R., & Jaggers, K. (2013). Polity IV project: Political regime characteristics and transitions, 1800-2012. Center for Systemic Peace.

Mayda, A.M. (2006). Who is against Immigration? A cross-country investigation of individual attitudes towards immigrants *The Review of Economics and Statistics* 88(3): 510-530.

Mayda, A.M. & G. Peri & W. Steingress (2018). The political impact of immigration: Evidence from the United States. NBER Working Paper No. 24510.

McLaren, M.L (2003). Preferences for the Exclusion of Migrants. *Social Forces*, 81(3): 909-936.

McHugh, M. C., & Frieze, I. H. (1997). The Measurement of Gender-Role Attitudes. A Review and Commentary. *Psychology of women quarterly* 21(1): 1-16.

Mendez, I. & I.M. Cutillas (2014). Has Immigration affected Spanish presidential elections results? *Journal of Population Economics* 27: 135-171.

Moaddel, M. & T. Azadarmaki (2002). The Worldviews of Islamic Publics: The Case of Egypt, Iran and Jordan. *Comparative Sociology* 1(3-4): 299-319

Montalvo, J.G., and Reynal-Querol, M. (2003). Religious polarization and economic development *Economics Letters*, 80(2):201–210.

Montalvo, J.G., and Reynal-Querol, M. (2005). Ethnic polarization, potential conflict, and civil wars. *American Economic Review*, 95(3):796–816.

Moriconi, S., Peri, G., and Turati, R. (2018). Skill of the Immigrants and Vote of the Natives: Immigration and Nationalism in European Elections 2007-2016. NBER Working Paper No. w25077.

Murard, E. (2017). Less Welfare or Fewer Foreigners? Immigrant Inflows and Public Opinion towards Redistribution and Migration Policy. *IZA DP* No. 10805

Murtin, F. & Wacziarg, R. (2014). The democratic transition. *Journal of Economic Growth* 19(2): 141-181

Myers, S. (2000). The impact of religious involvement on migration. *Social Forces* 79: 755-783

Nickel, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49(6):1417–1426.

Noland, M. (2005). Religion and Economic Performance. *World Development* 33(8): 1215-1232

Norris, P. & R. Inglehart (2011) Sacred and Secular. Religion and Politics Worldwide. Cambridge University Press.

Nyhan, B. & Jason Reifler (2010). When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior*, 42:303–300.

Orrenius, P. (2017). New Findings on the Fiscal Impact of Immigration in the United States Dallas FED Working Paper 1704.

Ortega, F., and Peri, G. (2014). Openness and income: The roles of trade and migration. *Journal of International Economics*, 92(2):231-251.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics* 37(2): 187-204.

Ottaviano, G., and Peri, G. (2006). The economic value of cultural diversity: Evidence from U.S. cities. *Journal of Economic Geography*, 6(1):9–44.

Ottaviano, G and G. Peri (2012). Rethinking the Effects of Immigration on Wages. *Journal of the European Economic Association*, 10(1):152-197

Otto, A.H. & M.F. Steinhardt (2014). Immigration and election outcomes - Evidence from city districts in Hamburg. *Regional Science and Urban Economics* 45: 67-79.

Özden, Ç., Parsons C. R., Schiff, M., and Walmsley, T. L. (2011). Where on earth is everybody? The evolution of global bilateral migration 1960-2000. *World Bank Economic Review*, 25(1):12–56.

Ozgen, C., Peters, C., Niebuhr, A., Nijkamp, P. and Poot, J. (2014). Does cultural diversity of migrants employees affect innovation? *International Migration Review*, 48(9):377–416.

Parrotta, P., Pozzoli, D., and Pytlikova, M. (2014). Does labor diversity affect firm productivity? *European Economic Review*, 66:144–179.

Pendakur, K. & R. Pendakur (2005). Ethnic Identity and the Labour Market. Working paper, Department of Economics, Simon Fraser University.

Peri, G. (2012). The effect of immigration on productivity: evidence from the U.S. States. *The Review of Economics and Statistics*, 94(1):348-358

Piketty, T. (1995). Social Mobility and Redistributive Politics. *Quarterly Journal of Economics* 110(3): 551–84

Poutvaara, P. & Steinhardt, M. (2018). Bitterness in Life and Attitudes Towards Immigration," *European Journal of Political Economy*, 55: 471-490.

Preacher, K.J. & R.C. MacCallum (2003). Reparing Tom Swift's Electric Factor Analysis Machine. *Understanding Statistics* 2(1): 13-43

Price, A. (2015). How national structures shape attitudes toward women's right to employment in the Middle East. *International Journal of Comparative Sociology* 56(6): 408–432

Putnam, R. D. (2000). Bowling alone: America's declining social capital. New York: Simon & Schuster.

Putterman, L., and Weill, D.N. (2010). Post-1500 population flows and the long-run determinants of economic growth and inequality. *The Quarterly Journal of Economics*, 125(4):1627–1682.

Rapoport, H. & S. Sardoschau & A. Silve, (2018). International migration and Cultural convergence. Manuscript.

Richards, A. & J. Waterbury & M. Cammett & I. Diwan (2014). A Political Economy of the Middle East. Third Edition. Boulder, Colorado, USA: Westview Press.

Rodrik, D. (2018). Populism and the Economics of Globalization. *Journal of International Business Policy* 1-22.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician* 39(1): 33-38.

Rosenbaum, P. R. (2002). Overt bias in observational studies. In Observational studies (pp. 71-104). Springer, New York, NY.

Ruggles, S., Genadek, K., Goeken, R., Grover, J., and Sobek M. (2015). Integrated Public Use Microdata Series: Version 6.0 [Machine-readable database]. Minneapolis: University of Minnesota.

Ruyssen, I. & S. Salomone (2018). Female Migration: a Way out of Discrimination? *Journal of Development Economics* 130: 224-241

Sadiqi, F. (2014). Women's NGOs and the Struggle for Democracy in Morocco. in: Galia Golan and Walid Salem (eds.) Non-State Actors in the Middle East: Factors for Peace and Democracy. University of California at Los Angeles Center for Middle East Development. New York, NY, USA: Rutledge.

Safadi, R.& S. Neaime (2017). Syria: The Painful Transition Towards Democracy. in: Ibrahim Elbadawi and Samir Makdisi (eds.) Democratic Transitions in the Arab World, Cambridge, UK: Cambridge University Press.

Santos Silva, J. M. C., and Tenreyro S. (2006). "The Log of Gravity". *Review of Economics and Statistics*, 88(4):641–658.

Scheve, K.F. & M.J. Slaughter (2001). Labor market competition and individual preferences over immigration policy. *Review of Economics and Statistics* 83(1): 133-145.

Senik, C., Stichnoth, H. & Van der Straeten, K. (2009). Immigration and natives' attitudes towards the welfare state: evidence from the European Social Survey. *Social indicators research* 91(3): 345-370.

Sequeira, S., N. Nunn and N. Qian (2017). Migrants and the making of America: the short- and long-run effects of immigration during the age of mass migration. NBER Working Paper No. 23289

Shenkar, O. (2012). Cultural Distance Revisited: Towards a more rigorous conceptualization and measurement. *Journal of International Business Studies* 43(1): 1-11

Shweder, R.A. & R.A. Levine (1984). Culture Theory: Essays on Mind, Self and Emotion. New York: Cambridge University Press.

Sianesi, B. (2004). An evaluation of the Swedish system of active labor market programs in the 1990s. *Review of Economics and Statistics* 86(1): 133-155.

Sinding Bentzen, J. (2019). Acts of God? Religiosity and natural disasters across subnational world districts. *The Economic Journal* 129(622): 2295-2321.

Solow, R. M. (1956). A contribution to the theory of economic growth. *The Quarterly Journal of Economics*, 70(1):65-94.

Spilimbergo, A. (2009). Democracy and foreign education. *The American Economic Review*, 99(1):528–543.

Spolaore E, and Wacziarg R. (2009). The diffusion of development. *Quarterly Journal of Economics*, 124(2):469–592.

Spolaore, E. & R. Wacziarg (2012). Long-Term Barriers to the International Diffusion of Innovations. In: J. Frankel and Ch. Pissarides (eds.), NBER International Seminar on Macroeconomics 2011, Chapter 1, 11-46.

Spolaore, E. & R. Wacziarg (2013). How Deep are the Roots of Economic Development? *Journal of Economic Literature* 51(2): 325-369

Spolaore E, and Wacziarg R. (2015). Ancestry, Language and Culture. In: Ginsburgh V., Weber S. (eds) The Palgrave Handbook of Economics and Language. Palgrave Macmillan, London.

Steinmayr, A. (2016). Exposure to Refugees and Voting for the Far Right:(Unexpected) results from Austria. IZA DP No. 9790.

Stetter, S. (2008). World Society and the Middle East: Reconstructions in Regional Politics. New York, NY, USA: Palgrave Macmillan.

Stock, J. H., and Yogo, M. (2002). Testing for weak instruments in linear IV regression. National Bureau of Economic Research. Technical Working Paper No. 284

Storesletten, K. (2000). Sustaining fiscal policy through immigration. *Journal of Political Economy*, 108(2):300–323.

Suedekum, J., Wolf, K., and Blien, U. (2014). Cultural diversity and local labour markets. *Regional Studies*, 48(1):173–191.

Tabellini, G. (2010). Culture and Institutions: Economic Development in the regions of Europe. *Journal of European Economic Association* 8(4): 677-716

Tabellini, M. (2018). Gifts of the Immigrants, Woes of the Natives: Lessons from the Age of Mass Migration Harvard Business School Working Papers 19-005, Harvard Business School, revised Jan 2019.

Trax, M., Brunow, S., and Suedekum, J. (2015). Cultural diversity and plant-level productivity. *Regional Science and Urban Economics*, 53:85-96

Tuccio, M., & Wahba, J. (2015). Can I have permission to leave the house? Return migration and the transfer of gender norms. *IZA Discussion Paper* No.9216

Tuccio, M. & J. Wahba, & B. Hamdouch (2016). International Migration: Driver of Political and Social Change? IZA DP No. 9794

Tuccio, M., Wahba, J., & Hamdouch, B. (2019). International migration as a driver of political and social change: evidence from Morocco. *Journal of Population Economics* 32(4): 1171–1203.

Tzannatos, Z. (1999). Women and Labor Market Changes in the Global Economy: Growth Helps, Inequalities Hurt and Public Policy Matters. *World Development* 27 (3): 551- 569

UN (2005). Household Sample Surveys in Developing and Transition Countries. United Nations : New York.

U.N. Population Division. (2013). World population prospects: The 2013 revision.

UN (2015). Human Development Report 2015: Work for Human Development. United Nations : New York.

United Nations, Department of Economic and Social Affairs (2017). Trends in International Migrant Stock: Migrants by Destination and Origin. Data available online.

Valette, J. (2018). Do migrants transfer productive knowledge back to their origin countries? *The Journal of Development Studies*, 54(9):1637-1656.

Vasilakis, C. (2017). Massive Migration and Elections: Evidence from the Refugee Crisis in Greece. *International Migration* 56(3): 28-43.

Weber, M. (1946). The Social Psychology of World Religions. In : H.H. Gerth and C. Wright Mills (eds) pp. 267-301, in From Max to Weber, 267-301. New York: Oxford University Press.

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "p< 0.05". *The American Statistician* 73: 1-19.

Wood, C.H. & C. Gibson, & L. Ribeiro, & P. Hamsho-Diaz (2010). Crime Victimization in Latin America and Intentions to Migrate to the United States. *International Migration Review* 44(1): 3-24

World Bank Group (2013). World development indicators 2013. World Bank Publications, 2013.

Wuthnow, R. (1998). After heaven: Spirituality in America since the 1950s. University of California Press.

Young, H.P. (2008). Social norms. The New Palgrave Dictionary of Economics, 2nd Edition

Zhao, Z. (2004). Using matching to estimate treatment effects: Data requirements, matching metrics, and Monte Carlo evidence. *Review of Economics and Statistics* 86(1): 91-107.

Zelinsky, W. (1971). The hypothesis of the mobility transition. *Geographical Review* 91(2): 219-249