

Migration Crisis in the Local News: Evidence from the French-Italian Border.

Silvia Peracchi,¹

Keywords: *Media slant, EU borders, immigration, diff-in-diff*

JEL classification: F22, L82, F50

Migration Crisis in the Local News: Evidence from the French-Italian Border.

Silvia Peracchi, ¹

Abstract

This paper examines how undocumented migrant displacements affect local news markets and the political economy of EU internal borders. I focus on a policy that took place in June 2015, in which French authorities introduced militarized controls at their borders with Italy, to return waves of irregular border-crossing migrants transiting from Italy to Italian lands. Natives' exposure to resettled migrants varied across Italian municipalities, depending on their proximity to push-back areas. I exploit this quasi-experimental setting in a diff-in-diff framework and compile novel text and count data from local news in Liguria, Italy. Results show that migration coverage declined with distance from the border after the push-backs. In contrast, anti-immigrant discourse intensified away from the events and weakened at the border. Exploring further this framing dimension, I observe readers' demand to be closely associated with local news discourse and voting preferences to broadly follow anti-immigrant slant in the news.

Acknowledgements

I am thankful to M. Beine, Q. David, M. Fernandez Sanchez, C. Mahe, E. Murard, J. F. Maystadt, R. Poldermans, K. Tatsiramos, P. Van Kerm, S. Zana for precious comments and suggestions. I also wish to thank the participants and organizers of 12th European Meeting of the Urban Economics Association, the 2022 Seminar series on the Economics of Migration; the 15th RGS Doctoral Conference in Economics; the 2022 LISER Workshop on Globalization, Attitudes and Populism; the Sixth International Conference on Understanding Voluntary and Forced Migration; the International Conference of European Studies. This research project was part of the PhD thesis of the author, which was supported by the Luxembourg National Research Fund (FNR) (PRIDE15/10949242*). All outputs of this study are the sole responsibility of the author.

1. Introduction

Over the past decade, Europe has experienced an unprecedented rise in undocumented migration, with a remarkable peak between 2014 and 2016. Italy, in particular, saw a surge in arrivals of migrants and asylum seekers, departing from Libyan coasts on the so-called “Central Mediterranean Route.” In 2014 and 2015, arrivals along this route reached 170 thousand and 150 thousand individuals, respectively (European Border and Coast Guard Agency). These events intensified the public debate both at the national and EU levels, leading to viral attention in the media discourse (Chouliaraki et al., 2017).² Several scholars attributed the widespread success of populist, anti-immigrant parties in various EU countries to the direct exposure of natives to the presence of, or proximity to, migrant inflows (Campo et al., 2021; Edo et al., 2019; Otto and Steinhardt, 2014). On the one hand, media have been identified as crucial in shaping and spreading attitudes and expressed political preferences (Bratti et al., 2020; DellaVigna and Kaplan, 2007; Djourelouva, 2023; Gentzkow, 2006). On the other hand, media companies do not randomly select their messages to convey. News firms, in particular, study their competition and the beliefs of their readership in order to decide on the supply of information (Mullainathan and Shleifer, 2005).

In this study, I examine the effects of the 2010s migration crisis on the formation of local news, focusing on two key questions: How do border controls and resulting migrant displacement influence the characteristics of local news? Are these news patterns reflected the local political economy? In particular, I explore an exogenous shock in the presence of migrants at the Italian border. In June 2015, French authorities introduced militarized border controls at their border with Italy. Once the policy was introduced, French authorities inspected the migrants encountered along the entry points to France and resettled those without a permit to stay in France to the Italian lands at the border (specifically, in the territory of the town of Ventimiglia). Figure 1 exemplifies the geographic context, illustrating how the Alps and Apennines shape the border and

²For a comprehensive review of the refugee crisis, with a focus on the reception policies from EU members, one can refer to Dustmann et al. (2017).

make coastal areas more accessible. As a complement, an additional map realized by NGO La Cimade in Figure A.1 helps visualize the push-back dynamics.

Figure 1. Physical maps of France and Liguria, Italy



Notes: A physical map showing the north-western boundaries of Italy and the study region. Liguria and its provinces are filled in color (yellow, pink, blue and green). At its north, Piemonte appears delineated in black.

On the left of these two regions is France. National boundaries are highlighted with a thicker black line. Ventimiglia's location is marked in red, as well as Bardonecchia and Briançon. Source: OpenStreetMap (2023) and Google Terrain, accessed from QGIS Software: QGIS.org, on June 2024.

Before the border controls took place, many migrants located in Italy in precarious circumstances had taken advantage of this gateway to France to reach other European destination countries. As described in Campo et al. (2021), undocumented migrants reaching Southern Italy from Libya were subject to an allocation scheme that was reinforced in 2014. However, many intended to continue traveling, bypassing or interrupting the bureaucratic process of status regularization in Italy (Capitani, 2018; Colombeau, 2020).³ Although similar dynamics occurred along other gateways away from the coast, they included more impervious areas. One example is Bardonecchia (see Figure 1 and A.1).⁴

In this study, I exploit the geography of these border restrictions and focus on the municipalities of two

³For further information on how asylum seekers sorted in Europe, see also Aksoy and Poutvaara (2021).

⁴Additionally, gateways bordering other countries included, for instance, Como and Brennero (Capitani, 2018).

key provinces situated in the vicinity of the border: Imperia, the closest, and its adjacent province, Savona (highlighted in yellow and pink, respectively, in Figure 1). Whether locals directly experienced the presence of resettled migrants depends (negatively) on how connected they are to the border, which in turn is a function of commuting time to get there. Therefore, I employ the inverse of commuting time to the border to proxy the degree to which natives were exposed to the migrant push-backs. I build a municipality-month-year panel, which aggregates news information from a rich collection of about forty thousand migration-related articles from local news outlets. An interesting aspect of Local news is that it can reflect local differences in the salience of specific events, even in a context where migration is a central theme at the national level.⁵ The baseline framework is a continuous difference-in-difference model, which interacts border proximity with a dummy for the period after June 2015. I also conduct sensitivity checks to overcome the possible limitations associated with the TWFE estimators in the diff-in-diff framework (Callaway et al., 2021). Additionally, I perform a placebo replacing actual with random distances to other Italian cities.

There are two key outcome variables of interest: i) coverage of migration, representing the intensity of migration discourse and ii) A directed measure of "slant" or news-partisanship towards an anti-immigrant ideology, capturing the degree to which news narratives align with an anti-immigrant tone in migration-related reporting.⁶ To measure anti-immigrant slant, I rely on a binary classification algorithm that predicts whether a local news article has an anti-immigrant inclination. Predictions are based on a comparison of the similarity of an article's text with that of anti-immigrant versus more positively leaning national news outlets. This procedure builds upon the seminal works of Groseclose and Milyo (2005), which detected slant in US newspapers by counting citations for think-thanks linked with Republicans vs Democrats; and Gentzkow and Shapiro (2010); Taddy (2013) who developed an index of speech similarity to US Congress for news outlets, using more sophisticated techniques.

The main findings can be summarized as follows. The backlog of migrants led to a notable prominence of migration in the local news discourse at the border. It caused a differential increase in news coverage that sharply decays with the decrease in intensity of exposure to the pushed-back migrants. The share of anti-immigrant news, on the contrary, is observed to get relatively higher in the municipalities further away, i.e., where natives were least connected to the migrant settlement, and lowered at the border. Additional evidence proposes a set of mechanisms to rationalize this finding. First, the demand for news is found to have a crucial role. I construct a proxy for the consumption of local news, and for their penetration against national news, as measured via internet traffic from Google Trends. Looking at these outcomes, I find traffic and

⁵Specifically, I consider local news published online, given the widespread internet availability and usage in the region. In the sampled period, the region of interest had a substantial share of internet users in the population, which is higher than the Italian average (as shown in Figure D.3's internet usage rates over time. The plot compares the region of interest to the whole country).

⁶We can consider these characteristics as covering aspects of *agenda-setting* and *framing* in the taxonomy of the communication literature by Scheufele and Tewksbury (2007).

penetration of local news grow relatively more in further areas, in the same line as for anti-immigrant slant. Additionally, the effect of distance to the border on traffic is augmented if the slant is more anti-immigrant. According to the framework of Mullainathan and Shleifer (2005), readers have biased beliefs and media firms can accommodate these beliefs by slanting the information towards them. Where readers' opinions are more heterogeneous, their demand pushes bias in the two opposite directions, increasing news accuracy overall. Conversely, where consumers have homogeneous beliefs, bias is uniformly reflected in the news. In this context, direct exposure of locals to migrant settlement at the border appears to increase opinion heterogeneity, as documented in voting patterns. The co-occurrence of both negative and positive responses would push anti-immigrant bias down in aggregate. However, this comes at the expense of the interest of local readers, which grows the most far from the border.

One key potential channel of this variability is the recorded presence of humanitarian aid at the border from NGOs like La Cimade, Oxfam, Caritas, and Doctors Without Borders, potentially facilitating or reflecting direct local contact with the emergency. Far from the border, municipalities lacked a direct contact element, and the negative responses appear to prevail.

The contribution of this study is threefold. First, it connects to the literature on native responses to immigration and in particular to empirical contributions finding grounds for the contact hypothesis (Allport, 1954; Pettigrew, 1998). This theory was first proposed in social psychology and establishes that recurring interactions between individuals of different groups may enhance their mutual acceptance, under some conditions. Alesina and Tabellini (2024) frame the views of destination countries on immigration along two intrinsic dimensions: an economic and a non-economic, cultural one. Within this second strand, immigration inflows have been often associated with a stronger right-wing backlash and conservative ideology (Dinas et al., 2019; Hangartner et al., 2019; Bratti et al., 2020; Edo et al., 2019; Halla et al., 2017; Otto and Steinhardt, 2014; Gessler et al., 2021). However, effects have also been found to be heterogeneous or nuanced (Barone et al., 2016; Dustmann et al., 2019; Altındağ and Kaushal, 2020; Mayda et al., 2016) and even reversed in some circumstances (Lonsky, 2021; Bursztny et al., 2024; Gallegos Torres, 2023). In these references, scholars have associated positive responses to the exposure to immigrants with the contact theory framework (Allport, 1954; Pettigrew, 1998).⁷ Among recent findings, Bursztny et al. (2024) observe long-term interaction to lower hostility towards Arab-Muslims in the US. Christ et al. (2014) highlight how local interactions may allow in-group individuals to adopt a more positive response to the out-group if a critical mass of them engages in more tolerant norms. Dustmann et al. (2019) find an urban-rural divide in electoral voting in reaction to refugee migration, pointing at the role of interactions with migrants to explain better responses in urban settings. Most closely to this paper, Steinmayr (2021) finds evidence of the contact hypothesis, by looking at the refugee crisis in the European context, zooming on municipalities in Upper Austria close to Germany. The author distinguishes those units that hosted asylum seekers and had frequent interaction with them

⁷Other examples of evidence supporting the contact hypothesis are: Schindler and Westcott (2021), for the UK, Mousa (2020) for Iraqi nationals.

and municipalities that only experienced their transit, finding better attitudes when recurring contact was established.

The main results in the present study are in line with the contact-theory hypothesis observed in these contributions but differ from existing research as here a novel approach is taken on measuring and focusing on information and discourse spreading, in the form of local news. This reveals to be a crucial aspect representing spatial patterns in local narratives, which can both reflect and even anticipate other manifestations of ideology, such as voting or protesting. For example, Campante et al. (2018) showed how the spread of broadband internet in Italy from the '90s to the early 2000s was decisive for a large and coordinated mobilization of dissatisfied citizens, that went from exiting the electorate to voicing their preferences via voting for the new, Internet-born, Five-Star Movement. In this sense, researchers studying the contents of online information can capture almost simultaneously how politically relevant patterns evolve.

Second, this study connects to research investigating the persuasive power of news (Prat and Strömberg, 2013), but here the question is reversed and the production of news is allowed to be an endogenous choice. This literature has considered several means of communication: radio (DellaVigna et al., 2014; Strömberg, 2004), printed press (Gentzkow, 2009; Snyder Jr and Strömberg, 2010), TV (Campante et al., 2018; Facchini et al., 2017; Gentzkow, 2006) and the internet (Falck et al., 2014). More related work has focused on migration-specific sentiments (Benesch et al., 2019; Iannelli et al., 2021; Battiston, 2022; Djourelouva, 2023; Keita et al., 2023), and other connected outcomes (de Coulon et al., 2016; Battiston, 2022).

Finally, while relying on a novel context and original data, this work broadly connects to a body of studies in political economy on the drivers of the production of information; see, for instance, the theoretical framework of Gentzkow et al. (2015), the role of mergers in Anderson and McLaren (2012), Gentzkow and Shapiro (2010)'s study on newspapers' incentives to slant the content of news; George and Waldfogel (2006) looking at national vs local news penetration and especially Gentzkow et al. (2019a,b), which introduce the use of natural language processing to analyze the dynamics of politically charged news content.

While analyzing a specific geographic context, the present study importantly generalizes to areas that experience border tensions in the context of migration. These phenomena are non-negligible: in the European Union's context, internal borders have been the object of political tensions between national governments on several occasions and this study sheds light on the political economy of these areas. It urges for better border governance and cooperation to ultimately prevent the rise of extremist views in areas that receive only partial and indirect information around migration. The rest of the paper proceeds as follows. Section 2 explores the setting and context of the study. Section 3 details the data and measurement. Section 4 describes the empirical strategy, while the core analysis follows in section 5. Section 6 addresses the potential mechanisms, continuing with a set of robustness tests. Finally, the last section provides concluding remarks.

2. Context

2.1. Italy's Migrant Crisis and the Border Controls

During the 2010s, thousands of undocumented migrants landed in Italy with the intention of rejoining family members and contacts in other European countries. This surge was exacerbated by several political contingencies in developing countries, a key event being the “Arab Spring” in the MENA regions.⁸ In 2014, the Italian government enacted an allocation policy to strengthen a pre-existing, insufficient reception system for asylum migration. However, several migrants attempted to continue the journey instead of having their status regularized (Capitani, 2018; Colombeau, 2020).⁹ A typical gateway employed by migrants was located on the coastal border between France and Italy, in the north-western region of Liguria.¹⁰ The geography of Liguria, a short strip of land, extends east and west (Figure 2 displays its localization). On its northern boundaries are the Maritime Alps and Ligurian Apennines, while the Mediterranean Sea is on the south. Mountains delineate the French border region as well, making the coastal border town of Ventimiglia a crucial point of contact with France.

Following peaks in arrivals in southern Italy, the French local authorities introduced a militaristic border control in June 2015. This was done to push unauthorized migrants who attempted to reach France through the coastal gateway back to Italy, specifically to the border town of Ventimiglia. In the immediate wake of the policy’s implementation, the local charity Caritas Intemelio recorded two hundred migrants seeking refuge on the streets; nonetheless, these numbers continued to rise during the summer months.

In a number of reports, the location of Ventimiglia was dubbed as “Italian Calais”. The name evoked a similar bottleneck situation and precarious living conditions that were endured by those migrants trying to enter the UK by sea from France. The topography of the French-Italian border made it difficult for displaced people to locate easy alternative ways to cross the border. Although some immigrants thought of taking a different route to the north across the mountains, this option was relatively unpopular because it was arduous and perilous (Welander, 2017).

Due to the occurrence of terrorist attacks, in late 2015 France declared a state of emergency, thus officially establishing the border closure. After a second peak in May 2016, several NGOs came in support of the emergency, and the Italian provincial authority opened a temporary Red Cross reception center. In the same period, the Italian central government established a bus system to transport some of the migrants from Ventimiglia back to the major migration hubs in southern Italy. Nevertheless, Ventimiglia’s border only saw a significant decline in arrivals in late 2017 as a result of a set of policies at the EU and national level, which effectively decreased the number of migrants attempting the Central Mediterranean Route. In particular,

⁸The International Organization for Migration (IOM) reports 62692 undocumented migrants entering Italy by sea in 2011, compared with 4406 and 9573 respectively in 2010 and 2009.

⁹Once landed in Southern Italy, migrants could submit an asylum application. As importantly mentioned in Campo et al. (2021), this procedure is remarkably lengthy.

¹⁰Others were Bardonecchia, Como, Brennero (Capitani, 2018).

Italy had signed a Memorandum of Understanding with Libya in February 2017 (which was later extended) with the intention of cooperating with Libyan coast guard authorities to prevent the boat lifts. Sourced by NGO La Cimade, Figure A.1 is a good summary of the timeline and geography of the main events.

Structured data are lacking on the presence of migrants in Ventimiglia. However, in 2017, NGO Caritas Intemelia collected a count of meals distributed to migrants. Figure A.2 in the Appendix reports a peak of 12,500 migrants transiting in the area in July 2017. For comparison, Ventimiglia’s population was approximately twenty-four thousand residents in 2019 (Italian National Institute of Statistics [ISTAT]). Despite measures being taken, the situation persisted over time. With some seasonality, migrant arrivals in Ventimiglia were also observed in 2018 and 2019. Appendix section Appendix C elaborates the political context of Italy in greater detail.

Figure 2. Area of interest, Liguria



Notes: A reference map of the region of interest displays a close-up view of Liguria on the left side. Ventimiglia, the border town of interest, is located in the west. The capital cities of each province in the region are labeled, including Imperia, Savona, Genova (the regional capital), and La Spezia. Imperia’s province is shown in blue, and Savona’s province is in green. Provinces in grey are not part of the sample. On the right side, there is a map of Italy with Liguria highlighted in blue.

3. Data and Measurement

The final data is a monthly panel of municipalities in the provinces of interest, for the periods between 2012 and 2019. Municipalities are assigned several socio-demographic characteristics, mainly sourced by ISTAT (details follow), and linked to the key news outcomes of interest. To produce news-based outcomes, I collect a rich corpus of news articles from online local outlets. More details on the data collection and measurement follow in next paragraphs and in the Appendix section Appendix D, and Appendix E.

3.1. News Corpus

News data are based on a novel corpus of local news articles for the sample of interest. For the collection of articles, I proceed in two steps: i) First, I identify which local online newspapers operated in the areas and time of interest; ii) Then, I extract their text and related information.

In the first step, I select all local news sources that could be located somewhere within the region and time period of interest. Importantly, this also includes outlets that appeared after the policy went into effect. This way, the data can be representative of news for all periods, which includes information from newly appearing sources. I initially list the names of local outlets by extensively searching on Google for those specific keywords that evoke localized news information (e.g. toponyms, such as names of bigger towns or "Riviera", plus "News"). I expand the initial list with a Google Trends feature that displays correlated searches. This tool indicates additional outlets that Google classifies as similar to the initial list, for the area of interest, where similarity is based on the users' internet searches.¹¹ Data inaccessibility prevents access to a minority of sources. In the Appendix, Table D.1 lists the identified outlets and a few that could not be tracked, adding details on their availability and creation dates. For sources that were partially or fully unavailable but whose names could be identified, the missing Google traffic amounts to a yearly average of 6.5% of the traffic for one popular source in the sample (Sanremonews).¹²

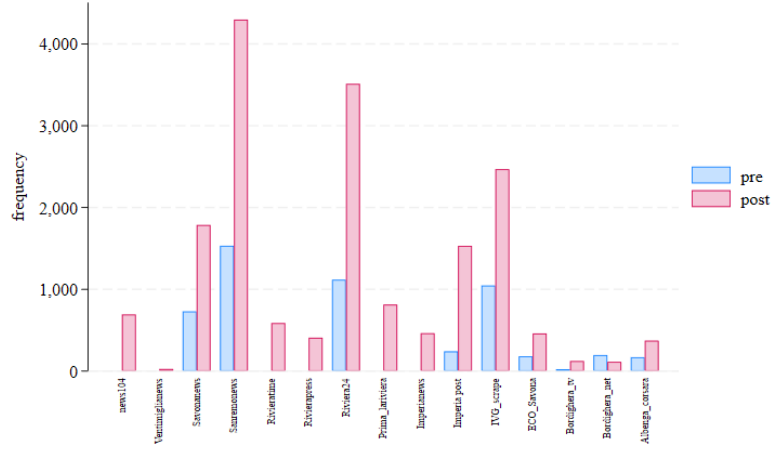
Once I identify the newspaper sources, I extract information from each local news website. To retrieve the articles, I adopt specific algorithms to extrapolate text from the articles in the websites' archive section or, alternatively, their search tool. The code differs for each news source, but involves 2 steps: first I gather the full list of news items of interest and their associated URLs, and then I loop over these URLs and extract the individual contents. I consider news items if they contain variants of the following keywords: "migrant," "immigrant," "foreigner," "non-EU citizen" (It., "extracomunitario"), or "displaced person" (It., "profugo"). Additionally, with a separate script, I extrapolate the overall news' count (including articles not related to migration) in the outlet per month, for each website. In sum, I retrieve 42,762 articles from 2012, coming from a total of fifteen distinct news sources.

Figure 3 displays the names and counts of news outlets in the sample, with a distinction in the metric using blue for before June 2015 and red for after June 2015.

¹¹For example, searching in Google Trends for the online outlet *Sanremonews*, would return correlated searches from users in the region for *Riviera24* and *Prima La Riviera*, which are outlets of interest as well.

¹²This number is retrieved from the data graphed in Appendix Figure D.1.

Figure 3. Sample collection, text corpus by news source



Notes: Distribution plot of the news in the sample, by news outlet. Blue bars represent data for the pre-treatment period while bars in red represent post-treatment frequencies.

In an additional cleaning step, I look at articles collected by matching the keyword “foreigner”. If those articles do not contain any other migration keyword, they might capture aspects different from migration (for example tourism). I analyze this group of articles through a topic-discovering method called Latent Dirichlet Allocation (LDA) (Blei et al., 2003; Hornik and Grün, 2011). With this procedure, I identify and drop a subset of around 10 thousand articles that are assigned topics unrelated to migration.¹³

3.2. News Data Measurement

Although online news is mainly accessible without any geographic boundary, local news firms tend to serve a particular area by tailoring their content and advertising revenues to the readers within their coverage. This implies that the geographic distribution of each local outlet actually varies. Therefore, to determine the areas that each local newspaper serves, it is crucial to understand their geographic target. With online news, we cannot count the number of newspapers sold in a neighborhood. The exact intensity of coverage of each article is also unobserved. To circumvent this limitation, I proxy this information by obtaining monthly internet traffic towards each news source at municipality level, as measured via Google Trends. Specifically, Google Trends tracks the searches for the outlets’ names over time and by municipality, reporting the following measures:

- i) $rank1_{sm\tau}$ is a traffic rating for source s by municipality m , at time-interval $\tau \in \{2010-2015; 2015-2020\}$,

¹³Details on the procedure are found in the Appendix section Appendix E.1.

where traffic information from Google is provided on a scale ranging from 1 to 100.¹⁴ Google Trends censors information for municipalities where traffic is too low. To limit the missing values resulting from this censoring, I select two wide time windows: i): January 2010- January 2015 and June 2015-January 2020, representing Google traffic at pre and post-border closure respectively.

To give an example, *Sanremonews* was mostly searched in the municipalities of Taggia (with a Google Trends rating of 100) and Sanremo (with a Google Trends rating of 75), in the period between January 2010 and January 2015. Figures D.5 and D.6 in the Appendix show Google Trends search results for the sources with the highest search traffic, in the two time ranges.

ii) $rank2_{st}$ is a rating of source s at month-year t . This measure doesn't vary by municipality but captures time variability at a higher frequency.¹⁵ The index appears again on a scale of 1-100. Importantly, if news sources display no traffic, their weight is zero in the baseline analysis. Figure D.7 in the Appendix shows how sources' online traffic fares over time, displaying some substantial variation across sources.

From this information, I construct the two main news outcomes as follows. News coverage of migration is a log count of articles that are matched to a municipality m in a given month-year t . This is the main outcome of interest to explore the intensity of migration discourse in local news.

$$\text{Coverage}_{mt} = \log\left(\sum_i^{N_t} \mathbf{1}_{\{i \in m, \tau\}}\right) \quad (1)$$

This measure sums articles i that are published in t and are assigned to municipality m if $rank1_{smt} > 0$.

The baseline measure of anti-immigrant slant, for municipality m at time t , is a weighted proportion of articles i , associated with municipality m from the Google Trends rating, that are predicted to be anti-immigrant:

$$\text{Anti-immigrant slant}_{mt} = \sum_{s \in S} w_{smt} \frac{1}{N_{smt}} \sum_{i \in N_{smt}} \mathbf{1}_{\{i \in s, m, \tau \& \hat{P}\{i = \text{anti-im}\} > 50\%\}} \quad (2)$$

Where w_{smt} varies by source, municipality and time. $\hat{P}\{i = \text{anti-im}\}$ is the predicted probability of article i being anti-immigrant. The indicator function in Equation 2 is one if the article is classified as anti-immigrant, i.e., if $\hat{P}\{i = \text{anti-im}\}$ is above 50%. I compute the proportion of anti-immigrant news at the source level and then aggregate them with a weighted sum. Weights are normalized to add up to 1 for every panel-month-year unit. More precisely: the *weight* w_{smt} of article i , published in source s , for municipality m , month-year t

¹⁴The platform normalizes search results by location and time range to avoid artificially giving too much weight to areas with overall high search volume. The resulting scaled outcome is a measure relative to all searches on all topics, according to Google Support references: <https://support.google.com/trends/answer/4365533?hl=en>. Retrieved in October 2021.

¹⁵To allow comparability of the ratings across sources, these indices are retrieved by using a common benchmark source for reference (i.e. outlet IVG, which reaches the highest relative traffic in the series period).

equals $w_{smt} = \frac{rank1_{sm\tau} * rank2_{st}}{\sum_{i \in m, t} (rank1_{sm\tau} * rank2_{st})}$. This gives more weight to sources with higher Google Trends traffic in that municipality and time period. As an example, suppose that, in a particular month, municipality m received news from 2 sources, L and R. Source L had zero anti-immigrant articles, while source R had 70% of its migrant articles predicted as anti-immigrant. If the traffic of source L is very high and that of R is very low, then the formula in Equation 2 will have a small value, because most readers read source L which bears no anti-immigrant content. If instead source L has close to zero internet traffic and source R has a high traffic level. Then the value of anti-immigrant slant in the equation would be high because users in the area mostly read the anti-immigrant content of source R.

Importantly, there’s a substantial variability in the degree of traffic across sources. Note that traffic is measured both before and after the policy occurrence, implying that its variation could be a result of the border closure. For this reason, I interpret this baseline measure of anti-immigrant slant as a demand-weighted measure. In the extension section, I explore alternative measures that zoom into the role of traffic as a separate outcome.

As mentioned, Google Trends censors locations below a certain level of search traffic. I impute values to missing municipalities using information from complete cases in near neighbors, defined using commuting statistics (Italian Census 2011, ISTAT).¹⁶ Further details are presented in Appendix section Appendix D.2.

The next section provides a detailed description of text-analysis techniques to obtain probability $\hat{P}\{i = \text{anti-im}\}$.

3.3. Measuring anti-immigrant speech

An article’s text is composed of its title, subtitle, and full body content. In addition to the text, the corpus of articles also includes their metadata: the author’s name (when provided), the article’s date, and tags (if available). I clean the text of sampled articles with common pre-processing practices.¹⁷ I adopt a *bag-of-words* framework for text analysis.¹⁸ This procedure relies on a matrix that is created to provide term

¹⁶A robustness test confirms the main findings without this imputation step.

¹⁷This involves the removal of punctuation, numbers, and *stop words*, i.e. words that are frequent in texts and serve as building blocks for sentences but do not add to the sentence content (e.g. *and, in, of, with, to, .*). I stemmed words to remove their endings (Bouchet-Valat, 2020). Uppercase (proper) names are removed (except if at the beginning of a sentence), and all words are set to lowercase.

¹⁸This technique represents a text document as an unordered combination of its terms appearances. The simplest version of this method employs individual words and their frequencies. Information on the words’ position in the sentence is discarded, meaning that the method relies only on the comparison of co-occurrences of terms in the documents. A straightforward extension, adopted here, replaces single words with pairs (bi-grams) and triplets of words (tri-grams). To some extent, this complication allows tracking word combinations that better represent a sentence’s meaning. See Gentzkow et al. (2019b) for a detailed review of these and alternative methodologies.

frequencies by documents; this matrix is known as *document-term matrix*, where documents are in rows i and terms (bi-grams and tri-grams, or n-grams more generally) are in columns j . The frequency of n-gram j in document i is specified as tf_{ij} =*term frequency*. For n-gram j , its *Inverse document frequency*= idf_{ij} is defined as the log of 1 over the share of documents containing n-gram j . The product of the term frequency and the inverse document frequency is termed $tf-idf_{ij}$ and is the metric used here. The advantage of this composite measure is that, unlike simple counts or frequencies, it avoids overweighting extremely rare or extremely common n-grams.

I feed the pre-processed text into a supervised machine-learning model of binary classification, which assigns a dummy label to each article. As mentioned in the previous section, the label takes a value of one if an article is classified as anti-immigrant and zero otherwise. To obtain the labels, I train a model on a different dataset, that is gathered from existing data with pre-classified labels. Note that the final measure is a *directed* media bias or *slant* toward right-leaning, anti-immigrant speech. Throughout this analysis, media bias and slant tend to be used interchangeably, in line with the definitions of Gentzkow et al. (2015) for media bias—"systematic differences in the mapping from facts to news reports [...] which tend to sway naive readers to the right or left on political issues" (page 3)—and of Gentzkow and Shapiro (2010) for media slant—"the frequency with which newspapers use language that would tend to sway readers to the right or to the left on political issues" (page 36).¹⁹ The rationale for focusing on right-leaning, anti-immigrant bias comes from the context of interest. As several authors point out (Campo et al., 2021; Bratti et al., 2020), anti-immigration positions in Italy emerged in two right-wing, populist parties *Lega* and *Fratelli d'Italia* primarily, though these parties integrated the main right-wing coalition during the years covered in this study. It is important to note that one could go further and distinguish the non-anti-immigrant articles between neutral and pro-immigrant. To do this, one would need to distinguish neutral from non-neutral news pieces. This task is not trivial, at least with the tools I adopt. Appendix section Appendix E.2 mentions a Yougov report, showing that audiences tend to disagree on the neutrality of national outlets. For this reason, I leave the assessment of neutrality for future research.

The classification procedure goes as follows. I gather a training sample containing articles whose labels are predetermined. I collect this sample from regional and national news outlets for which evidence exists that they express or do not have an anti-immigrant orientation. I base this evidence on a combination of survey and anecdotal knowledge, whose details are given in Appendix section Appendix E.2. News items classified as 1 in the training sample belong to newspapers *La Verità* and *Il Giornale*, while 0-labeled text is extracted from *L'Unità News* and *Il Secolo XIX*. To collect the training sample I proceed in the same way as I do to build the local news' corpus. From the term frequencies of this dataset, I estimate a model for the binary outcome, capturing the presence or lack of anti-immigrant slant. The parameters estimated in the

¹⁹Importantly, the notion of agenda-setting can also be understood as a form of bias (Prat and Strömberg, 2013). This paper opts for the terminology of Gentzkow et al. (2015) and Gentzkow and Shapiro (2010) and treats media bias and slant as substitutes.

model are then used to classify the sample of interest with local news items into the binary outcome variable.

I compare the performance of two alternative models: i) a regularized logistic regression with elastic-net penalization and cross-validated tuning parameters (Friedman et al., 2009); ii) an inverse regression model by Taddy (2013).²⁰ In an additional step, I also compare the model predictions with human annotations. The hand labeling is performed by *Amazon Mechanical Turk* annotators. Validation results suggest similar predictive performance for the two alternative models, with a slight preference for the use of Taddy (2013)’s measure, which is taken for the baseline regressions. Appendix sections Appendix E.2 and Appendix E.3 dig into the construction of this text-based measure in more detail and provide additional information on the validation procedure.

Figure A.3 explores predictions resulting from the classification. The picture reports frequent terms that appear in the articles with the lowest and highest predictive probabilities for anti-immigrant slant. Examples of expressions in blue at the bottom, which belong to text predicted as least likely anti-immigrant, are mild expressions to refer to migrants (e.g. “asylum seeker”) or terms related to migrants’ rights (e.g. “human rights,” “civil society”). Terms in the most likely anti-immigrant news (in red) relate, for example, to legal issues (e.g. “house arrest,” “controlled substance”), and expressions of concern on migrant flows (e.g. “illegal immigration”). Table B.1 accompanies the figure with a small list exemplifying local headlines in the sample. Finally, as mentioned in the paragraph above, classified articles are combined into a municipality-level metric that can be read as the average proportion of traffic for migration-related local news items classified as containing anti-immigrant language.

4. Empirical Strategy

This study adopts the following baseline model:

$$Y_{mpt} = \alpha_m + \delta_t + \beta_1 \text{InverseDistance}_m * \text{Post}_t + \Gamma X_{mpt} + \epsilon_{mpt} \quad (3)$$

where, for municipalities $m \in \{1, \dots, 132\}$ located in province $p \in IM, SV$ and month-year $t \in \{Jan2012, \dots, Dec2018\}$, we observe media outcomes Y_{mt} . Month-year and municipality fixed effects are represented by the terms δ_t and α_m , respectively. The coefficient of interest β_1 corresponds to the interaction term between the post-June-2015 dummy, and border proximity, measured with the inverse of commuting distance (in minutes) to the border area. In this sense, the degree of treatment represents the likelihood of individuals living in municipality m to be directly exposed to the border policy and migrants’ settlement. The degree of treatment decays with distance at an accelerating rate.²¹

²⁰Precisely, the model is cross-validated with 10 folds, i.e. it is estimated 10 times, each time including 9 out of 10 folds in the training sample and evaluating the model in the left-over subset.

²¹More precisely, the magnitude of this decay is proportional to the square of distance. For robustness, a set of extended results allows for different functional forms in the distance-based treatment.

X_{mt} is a vector of controls that vary with time and municipality. They're added gradually in the baseline estimates tables. To isolate the effect of distance from other geographic covariates, the first control added is an interaction between the post dummy and latitude, longitude and maximum altitude. In terms of news information, covariates involve the log of total news published in the local outlets, an average word count per outlet, and the count of outlets per municipality. Sourced by ISTAT, demographic variables are population (in logs), population density and the share of the population above 65 years of age, immigrant population stocks (in logs), as well as inflows and outflows. Additional covariates are the average taxable income per capita, extracted by the Home Office's tax registers, and province-level crime rates. Some specifications also control for the time distance from the administrative elections. The dates and years of local elections are not uniform in the sample, thus a variable is created by counting the months to the next election for each municipality. Further controls, only available cross-sectionally, are the unemployment rate, the rate of a highly educated population, altitude levels, and superficial area, all collected from the 2011 Census (sourced by ISTAT). These are included in the balance analysis as well as in a set of extended results. Table B.2 presents the summary statistics for these variables.

4.1. Identification and validity

To analyze the impact of migrant settlements at the western borders in Italy, this study concentrates on two provinces at the border region, namely, Imperia and Savona. These two areas are located respectively at the border and in its immediate vicinity. Such a restriction works in favor of internal validity, as it minimizes the influence of any other unobserved, co-occurring events related to the migration crisis in Italy, especially amidst the complexity of migration dynamics during this period. Another motive for relying on a localized setting is that I proxy the likelihood of locals to experience the migrants' settlement at the border with commuting distance from the border town of Ventimiglia. This likelihood fades away the more distant their municipality is from the border town. Limiting the sample geographically, therefore, both decreases the risk of the presence of confounders and assures that there's a strong variation in treatment exposure. The preferred treatment measure is the inverse of commuting time, which captures this dimension of proximity more accurately than aerial distance. In Figure A.4, data from the 2011 Italian Census shows the shares of individuals residing in the municipalities under study, who commute to the border or across, to work or study (y-axis), at different levels of distance to the border (in commuting time, x-axis). As the picture highlights, there's a strong prevalence of cross-border mobility next to the border. This visual result points to substantial variability in the degree of connections with the border for natives in the areas in the sample, especially in such localized regions.

In the main analysis, treatment is defined as the product between the degree of proximity and the post-June 2015 time interval. To better understand the treatment, it is important to highlight that both before and after the border policy, migrants were transiting the entire sample region. Once the border closed, however, the border area was suddenly exposed to groups of displaced people in precarious conditions, sheltering in and around the town. To ensure the validity of this approach for causal inference, municipalities

must be comparable along all varying distances. We need to assume that, in the absence of the border-related events, news outcomes in these municipalities would have followed parallel trends. To ensure this comparability, I first assess the balance of the sample, which shows whether some variables differ in levels, rather than trends. In the analysis, I then include various covariates in the regressions, conduct several robustness checks, and look into pre-policy trends in an event-study graph. I account for different province-specific time effects and finally propose a placebo that exploits distances to other areas in Italy, extracted randomly. Further details are provided in the following sections.

Another key aspect to discuss is the applicability of results to more general contexts, beyond these locations. Figures A.5 and A.6 compare the region in the sample with the distributions of several socio-demographic variables for all Italian provinces, estimated by kernel densities. The dashed vertical lines on the graphs indicate Imperia and Savona, in red and blue respectively. As shown in Figure A.5, Imperia has a relatively higher percentage of cross-border commuters. This aspect highlights that closer to the border, natives commute more to France, thus crossing the border regularly. This further justifies the choice of commuting distance for the treatment. Regarding the other variables, the two provinces are comparable to the rest of Italy. This holds in terms of total shares of commuters, employment rates, human capital distributions of the employed population, and shares of the immigrant population. In terms of the main immigrant origins, the areas of study present higher shares of the population from the American continent.

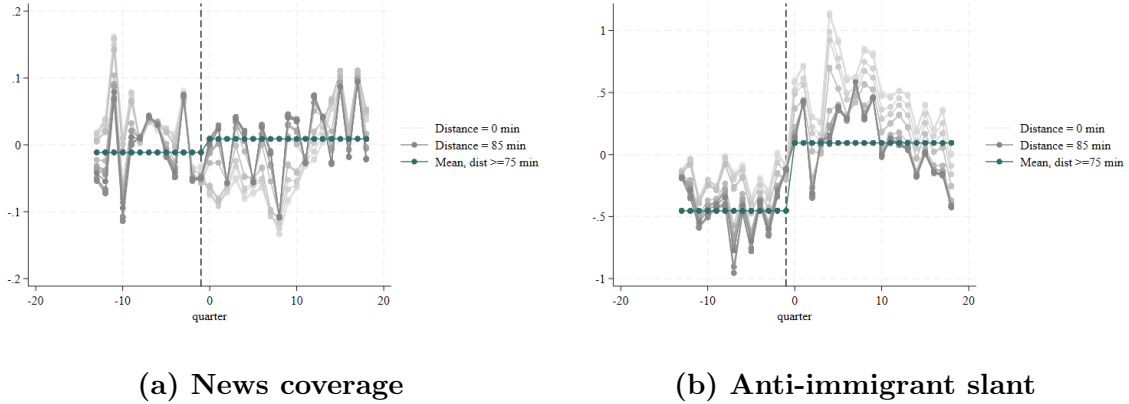
Figure A.6 depicts vote patterns from 2008 and 2013 for the Chamber of Deputies. With respect to the rest of Italy, the two provinces have comparable values of turnout and appear to lean more on the right of the political spectrum, with Imperia experiencing higher levels of right-wing incumbency than Savona.²² Besides these trends, a right-wing shift did appear in the area of study in the post-migration crisis elections of 2018. The right-wing coalition obtained on average 28.97% of votes for the Chamber of Deputies in 2013, and this increased to 44.94% in 2018. Left-wing coalition votes decreased from 23.89% in 2013 to 18.08%. This is important to keep in mind when generalizing results to other contexts. To better understand the role of some of these patterns, results section 5.3 investigates how the main effects depend on initial conditions, such as political preferences, as well as other socio-demographic characteristics.

Balance-test statistics appear in Figure A.7. The estimates result from a linear regression of one covariate on inverse distance, for the pre-policy subsample, accounting for time and province fixed effects. All variables are standardized to ease comparability. The table indicates that some differential in levels existed in the period before border closure for some variables. Population and mobility are higher closer to the border. Old-age shares are more frequent further from the border, as are the maximum altitude and the municipality's total area. Also the two outcomes of interest present slight but significant level differences. To circumvent potential issues arising from these pre-existing differences, observable characteristics that vary with time are included as controls in the model, while time-invariant covariates -unemployment, education, altitude and

²²For both provinces, 2013 constituted a relatively high success for the newly born anti-system, web-spread movement Movimento 5 Stelle (M5S).

initial political preferences- are further explored in an extended analysis, to investigate any heterogeneous effects depending on their levels. As it is crucial that level differences do not imply differential trends, several steps of the analysis explore these concerns. To start, Figure 4 plots the evolution of Y_{mt} means over time, by commuting distance (shown in quarters before and after the French border closure, represented as 0 in the x-axis). News coverage appears on the left and an anti-immigrant slant on the right. Means are computed for each time period and along spans of five minutes of commuting distance. The darker the line, the further the distance from the border. The picture delineates a clear switch in the link between distance and news coverage at the moment of the border events. A similar switch is also observable for the anti-immigrant slant. These descriptive patterns highlight a differential change in the news measures, resulting from the border closure. Besides the differential, a green line represents averages in the pre and post-periods, for observations situated at the highest distances (≥ 75 minutes). They highlight a sensible growth in coverage and a slight increase in anti-immigrant slant. This visual pattern will be assessed more structurally in the econometrics sections that follow.

Figure 4. Mean outcomes over time, by treatment



Notes: In this figure, means over time are plotted for the outcome variables Y_{mt} . The 0 level in the x-axis corresponds to the second quarter of 2015, which is the moment in which events took place.

5. Results

5.1. News Coverage

To what extent did the border controls to migrants translate into a sizeable media event? To answer this question, the top coefficients of Table 1 present the estimation results on news coverage - that is, the log count measure of migrant-specific news items.

In the table, controls are gradually added from left to right. News controls include the number of total news (in logs), as well as the average word count of the sampled articles. These are included in all specifications. Controls are added from Column 2 onwards. The second last column includes the count of local outlets present in the municipality m at month-year t . The count of outlets could be a function of the

formation of new demands and supplies of outlets as a response to the salient events at the border, therefore it may vary endogenously, but it is added in one specification because of its interest as a mechanism, as it allows exploring whether media effects are driven by changes in appearing sources versus changes occurring within sources. The last column adds province-specific time trends. This specification will be the preferred one to carry over in the extended results and robustness tests.

At the top of the panel, the interaction coefficients of interest are positive and significant, suggesting that news coverage of migration topics increased relatively more in the border’s vicinity. Regarding the size of the coefficient, we can interpret the output as follows. After the border events, increasing proximity to the border by 1 standard deviation from the average corresponds to an increase in coverage by a factor of $e^{0.03}$, or around 3%.²³ Because proximity is measured with the inverse of commuting distance, the effects of interest are sharply decaying with commuting distance from the border. The graph in Figure 5a fixes the pre-vs-post differentials estimated for the least treated municipalities as counterfactual (i.e. those located at least 75 minutes of commuting time from the border) and shows the effects of proximity, *ceteris paribus*. While coverage rose everywhere, the slope is steeper closer to the border.

To dig into the persistence of this impact, and to test for pre-existing linear trends, we can look at an event study specification in Figure 6 a. The coefficients relate to regression equation 4, where the effect is allowed to vary by quarter:

$$Y_{mpt} = \alpha_m + \sum_{\tau=Q1,2012}^{Q4,2019} \beta_t I[\tau = t] * InverseDistance_m + \Gamma X_{mpt} + \rho_{pt} + \epsilon_{mpt} \quad (4)$$

where m, p, t and ΓX_{mpt} are defined as in equation 3. The term $\rho_p * t$ captures province-specific time effects. The set of controls is equivalent to Column 6 (i.e., the richest set, excluding the outlet counts, to avoid controlling for a media-related outcome), and taking the quarter before the events (Q1 2015) as the reference period.

As Figure 6 a highlights, the effect of higher coverage closer to the border persists for several periods after the border events. Some fluctuation appears, with some spikes in some quarters of 2016, and some fading in the most recent periods. Besides these fluctuations, the other coefficients appear similar in size to the baseline results.

5.2. Anti-immigrant slant

Baseline results on the anti-immigrant slant are presented at the bottom of Table 1. Estimation results of Table 1, Column 6, indicate that, as a result of June 2015’s events, anti-immigrant slant in local media got relatively higher in areas less directly concerned, as located further away. An increase in proximity by 1 standard deviation from the average corresponds to a decrease in anti-immigrant slant by 0.084 standard deviations. Again, this effect can be better understood in terms of the pre-vs-post differential along commuting distance, as depicted in Figure 5b minutes away. The graph takes the furthest areas as counterfactual (defined

²³Border proximity is standardized to ease interpretation.

as being more than 75 minutes of commuting for the border). While anti-immigrant slant grows in the counterfactual, it grows less positively the closer to the border. The estimates indicate a flatten slope at around 10 minutes away and the slope direction reverses at the border (within 5 minutes). This result suggests that after the border policy, attitudes improved at the border and worsened with distance.

Figure 6 b shows results from the event study specification as detailed for the media coverage results. In the plot, pre-policy coefficients are not significantly different from zero and the post-policy ones have negative point estimates. While the coefficients on the right-hand side of June 2015 display some fluctuation in their significance, the effect appears relatively persistent, fading away in the latest periods. Several sensitivity analyses for baseline results are examined in Section 6.

Table 2 presents robustness checks with respect to some variations in the measurement of anti-immigrant slant.

In the first column of Table 2, the alternative classification method by Friedman et al. (2010) is employed.²⁴ In the baseline measure, a label of 1 is assigned to an article, if the predicted probabilities obtained in the classification model are higher than 0.5. The second and third columns employ alternative threshold probability scores of 0.55 and 0.45 to assign anti-immigrant slant labels to articles. Results from these specifications stay in line with the baseline, both in significance and size.

Proceeding to the right, an alternative specification directly considers probability scores, rather than labels, as the article-level outcome to aggregate at the municipality level. Results are presented for both the benchmark index and for the index based on Friedman et al. (2010), leading to similar results. The last column shows the results considering the size of anti-immigrant news traffic, controlling for coverage. This last column will be better explained in subsequent paragraphs. The coefficients are negative and significant, confirming what was found in the baseline analysis. Figure 7 complements these robustness tests by showing estimated coefficients from anti-immigrant slant classification based on probability matching. The variability results from 1000 iterations.²⁵ In the graph, one point in each vertical segment corresponds to the estimated effect based on one out of 1000 regressions (each regression includes a different iteration of the probability matching). When incorporating the classification uncertainty in this way, we find a similar visual pattern to

²⁴See also Friedman et al. (2009) detailing the R package employed to classify news according to this method.

²⁵According to this method, every article is assigned the 0-1 label, by drawing randomly from the discrete distribution $\{(1-\hat{P}\{i = \text{anti-im}\}), \hat{P}\{i = \text{anti-im}\}\}$. The draws are repeated a number of times M , 1000 in this case, and the average class across iterations of the same article converges to the predicted probability. While the baseline hard classification is the optimal decision rule to map predicted probabilities into labels in terms of accuracy, this method allows to capture and carry over the uncertainty that results from the classification. Suppose for article A, the anti-immigrant slant is predicted with a probability of 0.51, while for article B, the predicted probability is 0.99. With probability matching, article A will not be classified as anti-immigrant in almost half of the 1000 iterations, while article B will be classified as anti-immigrant almost always.

the baseline event study. This suggests that results are robust to several measurement variations.

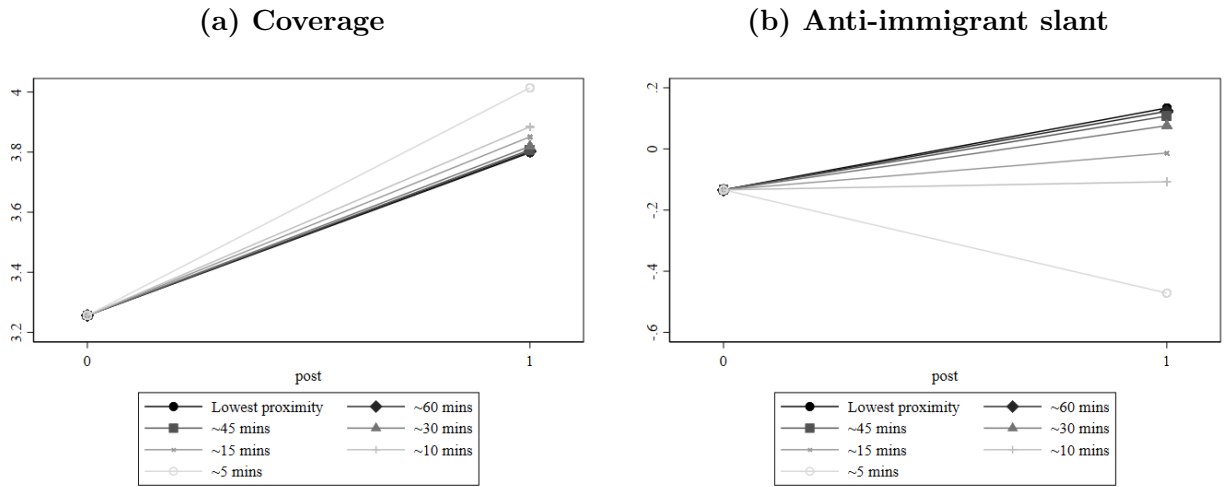
Is the magnitude of the effects comparable to existing estimates? Though literature is scarce, Branton and Dunaway (2009) look at a broadly similar relation between distance to the border and media outcomes, and focus on the U.S.-Mexico border in California. The authors correlate Californian news coverage of immigration and distance to the borders, for the period 2004-2005, and find that an extra ten miles of distance to the border relates to a 1.98% lower coverage of Latino immigrant news. Comparing the present study with this correlation, and assuming 10 miles can be traveled in 20 minutes, baseline results suggest a much steeper slope, that decays to comparable values from around 23 minutes away from the border onward. The authors also find distance to correlate negatively with coverage of negative aspects of immigration. By exploiting a quasi-experimental setting, the present study finds an opposite pattern, which suggests the importance of considering shocks in the salience of distance, besides the exploration of general tendencies. In the US setting, Gentzkow and Shapiro (2010) predict an 18 % loss in circulation of an average (national) newspaper if it were to deviate by 1 standard deviation from the profit-maximizing level of slant. Taking this reference as a crude benchmark, assuming that slant adjustments result from news firms maximizing profits, keeping slant values unchanged would decrease readers' interest by roughly 4.7% in the municipalities within the highest distance.

Table 1. Baseline results: News coverage and anti-immigrant slant

Dependent variable: Migrant news coverage (log)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Inverse distance (min)							
# Post	0.055*** (0.011)	0.041** (0.013)	0.038** (0.013)	0.038** (0.013)	0.038** (0.013)	0.038** (0.013)	0.030** (0.012)
Adjusted within R ²	0.04	0.05	0.05	0.05	0.05	0.05	0.22
$\hat{Y}Dist \geq 75$ min, PRE	3.261	3.252	3.252	3.252	3.253	3.253	3.256
$\hat{Y}Dist \geq 75$ min, POST	3.76	3.766	3.802	3.802	3.801	3.801	3.798
Dependent variable: Anti-immigrant slant							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Inverse distance (min)							
# Post	-0.256*** (0.059)	-0.091** (0.043)	-0.085* (0.046)	-0.085* (0.046)	-0.078* (0.046)	-0.083* (0.045)	-0.083** (0.036)
Observations	12636	12636	12129	12129	12129	12129	12129
Adjusted within R ²	0.06	0.20	0.21	0.21	0.22	0.23	0.08
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geo and News		Yes	Yes	Yes	Yes	Yes	Yes
Months to elections			Yes	Yes	Yes	Yes	Yes
Population controls			Yes	Yes	Yes	Yes	Yes
Migration controls				Yes	Yes	Yes	Yes
Income and crime					Yes	Yes	Yes
Tot. outlets						Yes	
Prov. level trends							Yes
$\hat{Y}Dist \geq 75$ min, PRE	.032	-.167	-.166	-.166	-.174	-.175	-.134
$\hat{Y}Dist \geq 75$ min, POST	.049	.197	.16	.16	.166	.168	.134

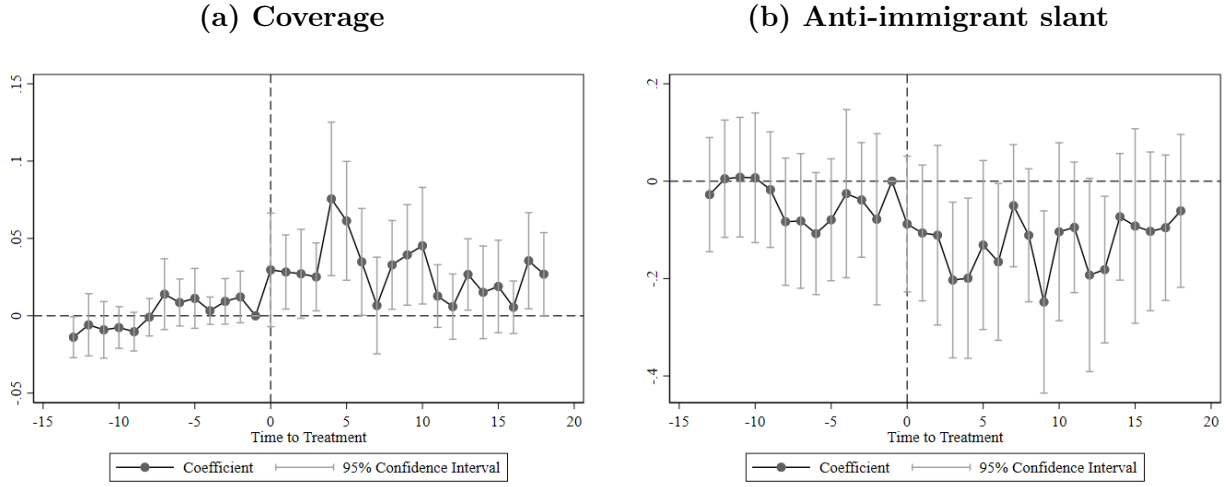
Notes: Baseline results for news coverage and anti-immigrant slant: results pertain to a reduced form diff-in-diff model outlined in Equation 3. The dependent variable is the log of news counts in the top-panel and a anti-immigrant slant at the bottom. Standard errors are two-way clustered at the municipality and month-year level. Controls are gradually added from left to right. Stars correspond to the following p-values: * $p < .10$, ** $p < .05$, *** $p < .001$.

Figure 5. Predicted outcome, by treatment



Notes: Average predicted outcomes, pre-vs post, taking the least treated observations (≥ 75 minutes from the border) as counterfactuals, in the pre and post periods. We fix the change in the counterfactual group and see how increasing proximity (*ceteris paribus*) affects the outcome levels.

Figure 6. Event study, effect over time.



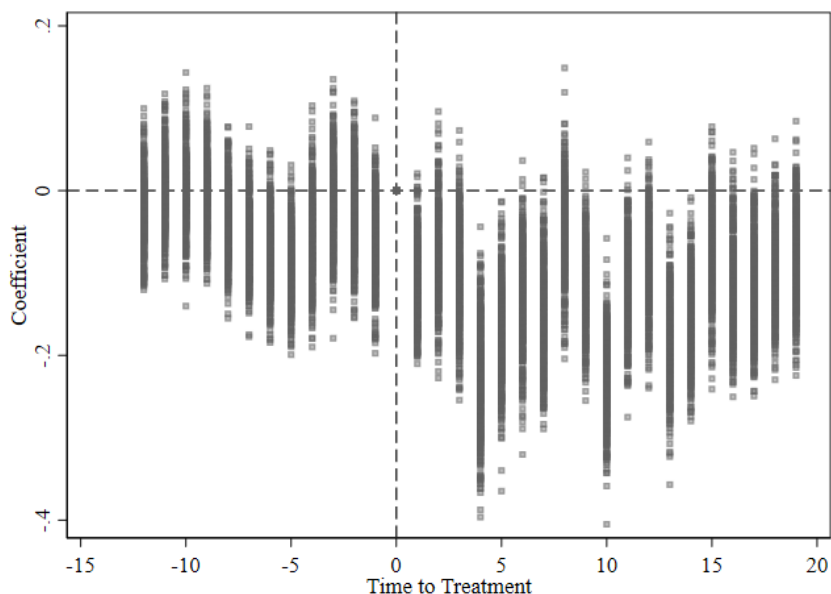
Notes: Coefficients plot from an event study specification. The outcome of interest is respectively news coverage in subfigure a, and anti-immigrant slant in b. These specifications control for province-level-time FEs. These inclusions allow to shut down the significance of some pre-policy coefficients, that is otherwise present for some period-specific coefficients. Estimations are based on twoway-clustered (by panel id and time) standard errors.

Table 2. Extended results: Anti-immigrant slant

	Dep. variable: <i>Slant</i>			Dep. variable: <i>Predicted probabilities</i>		Dep. variable: <i>Size</i>
	(1)	(2)	(3)	(4)	(5)	(6)
	Friedman et Al. 2010	Threshold 55	Threshold 45	Friedman et Al. 2010	Taddy (2013)	Anti-im traffic
Inverse distance (min)						
# Post	-0.079** (0.037)	-0.091** (0.039)	-0.086** (0.037)	-0.090** (0.039)	-0.098** (0.041)	-0.014** (0.006)
Observations	12129	12129	12129	12129	12129	12129
Adjusted within R ²	0.06	0.09	0.08	0.07	0.10	0.73
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Geo and News	Yes	Yes	Yes	Yes	Yes	Yes
Months to elections	Yes	Yes	Yes	Yes	Yes	Yes
Population controls	Yes	Yes	Yes	Yes	Yes	Yes
Migration controls	Yes	Yes	Yes	Yes	Yes	Yes
Income and crime	Yes	Yes	Yes	Yes	Yes	Yes
Tot. outlets						
Prov. level trends	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Extended results for news anti-immigrant slant : results pertain to a reduced form diff-in-diff model outlined in Equation 3. The set of controls is equivalent to the baseline's last column. Dependent variables vary by column and constitute a variation of the baseline slant index, except for the last column, which represents results on online traffic. Standard errors are two-way clustered at municipality and month-year level. Stars correspond to the following p-values: * $p < .10$, ** $p < .05$, *** $p < .001$.

Figure 7. Event study, slant effect over time, robustness to multiple predictions.



Notes: Coefficients plot from an event study specification based on 1000 regressions. The outcome of interest is anti-immigrant slant. In each regression, articles are classified based on probability matching. These specifications control for individual time and seasonal trends as for the baseline event-study graphs.

5.3. Heterogenous Results by Time Invariant Controls

In the baseline results, I find that a local shock in migrants' presence increased news coverage of migration in the most exposed areas. Conversely, the shock made anti-immigrant slant rise more in the municipalities least directly exposed, with the effect inverted at the border. In Table B.3, I explore the heterogeneity of these findings and analyze whether effects differ by the municipality's average education levels (based on 2011 shares of University graduates), initial incumbent ideology (coming from vote-shares for the left-wing coalition in the general elections of 2008), altitude levels (measured with maximum altitude), and initial unemployment rates (based on 2011 census data). Column 1 is included for benchmark and is equivalent to the rightmost of the baseline columns. The top of the table shows results on news coverage. Evidence suggests that the closest areas to the border receive more news coverage and that this effect is stronger in areas with higher altitudes (lower altitude municipalities can be conceived as urbanized, dense areas, given the region's morphology). Results on anti-immigrant slant are also heterogeneous along altitude levels. The negative effect of proximity to the border on slant is more pronounced in areas with higher altitudes, higher left-wing incumbency and higher education levels. Results are relatively homogeneous across the other characteristics considered.

5.4. Voting Patterns

How do the outcomes in local news match with aspects of the local political landscape? To answer this question, I explore how migrants' push-backs influenced voting patterns, comparing the periods before and after the border control policy.²⁶ I base this analysis on a first-differences specification, where differences are computed between the general election years 2018 and 2013.

The equation line is the following:

$$\Delta Y_{mc} = \beta_1 \text{InverseDistance}_m + \beta_2 \text{InverseDistance}_m * \text{Anti-immigrant slant}_m + \beta_3 \text{InverseDistance}_m * \text{Coverage}_m + \Gamma \Delta X_{mc} + \epsilon_{mc} \quad (5)$$

For municipalities $m \in \{1, \dots, M\}$, we observe election outcomes ΔY_{mc} , corresponding to the difference in vote shares for a party or coalition in 2018 with respect to 2013, for each of the parliament chambers c : {Senato, Camera}. The treatments of interest are a combination of proximity to the border (with corresponding coefficient β_1), and its interactions with the average from 2015 to the election year of media outcomes $\text{Anti-immigrant slant}_m$ and Coverage_m (with corresponding coefficients β_2 and β_3 , respectively). Table 3's specifications all control for population (in logs), population density the share of the population over 65, the log of immigrant stocks, income per capita, population inflows, outflows, and crime events (at province level), plus the geographic controls that are present in the baseline. The key covariates are standardized, to ease interpretation. Results on the shares of the two most explicitly anti-immigrant parties *Lega* and *Fratelli d'Italia* (FDI) are reported jointly in Column 1, and separately in Columns 2 and 3.

Far from the border, slant correlates positively with vote share changes for FDI, at the expense of the previous main party of the right-wing coalition PDL, and with voter turnout. Closer to the border areas, FDI gains success, more so if anti-immigrant slant is higher. However, the left-wing coalition also gains popularity in the vicinity of the border events, at a higher pace with higher coverage and slant levels. Coverage of migration appears to positively affect turnout and FDI shares in a similar fashion as anti-immigrant slant.

This evidence suggests that migrants' push-backs did not provoke a homogeneous positive response closer to the events, but rather sparked heterogeneous voting reactions. On one side they raised anti-immigrant voting, on the other side, the opposite ideology grew as well. Far from the border, news discourse favored anti-immigrant voting more uniformly. In light of the news-related findings, this heterogeneous response is a key mechanism to make slant less anti-immigrant in the vicinity of the pushbacks, relative to areas least directly involved, that had a rather uniform, relatively more negative response.

²⁶See Appendix Appendix C for a description of the Italian political context.

Table 3. Baseline results: Elections outcomes

Outcome: vote shares	Anti-immigration			Coalitions		Turnout	Other vote shares		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Lega, FDI	Lega	FDI	RW coalition	LW coalition	Turnout	PD	M5S	PDL, FI
Inverse dist	-0.026 (0.101)	-0.110 (0.096)	0.083** (0.037)	-0.121 (0.100)	-0.042 (0.063)	-0.031 (0.059)	-0.006 (0.049)	-0.068 (0.112)	-0.077 (0.084)
RW Slant	0.088 (0.079)	0.058 (0.075)	0.030** (0.013)	-0.065 (0.072)	0.063 (0.050)	0.100** (0.033)	0.053 (0.044)	0.009 (0.055)	-0.152*** (0.039)
Inverse dist #Average slant	0.001 (0.102)	-0.073 (0.099)	0.074** (0.029)	-0.046 (0.106)	0.196** (0.080)	-0.003 (0.066)	0.128* (0.067)	-0.071 (0.125)	-0.059 (0.090)
Inverse dist #Average coverage	0.013 (0.092)	-0.016 (0.091)	0.030 (0.022)	0.018 (0.118)	0.223** (0.080)	0.010 (0.065)	0.134** (0.067)	-0.036 (0.118)	-0.018 (0.107)
Coverage	0.067 (0.062)	0.031 (0.059)	0.036** (0.014)	-0.061 (0.058)	0.055 (0.043)	0.087** (0.032)	0.026 (0.037)	0.025 (0.052)	-0.132*** (0.035)
Observations	262	262	262	262	262	262	262	262	262
Adj. R squared	0.184	0.222	0.314	0.137	0.281	0.221	0.299	0.082	0.187
Mean Outcome	0.239	0.217	0.022	0.154	-0.065	-0.020	-0.066	-0.058	-0.082
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Results for voting outcomes. This table displays coefficients from a first differences model. The dependent variable constitutes the party or coalition's vote shares. Election outcomes from this table pertain to the general elections of 2013 and 2018. Standard errors are clustered at the municipality level. The set of controls is in first-differences and includes population (in logs), population density, over 65 population shares, logs of immigration stocks, income per capita, population inflows and outflows, and crime prevalence (in logs) at the province level, plus the same geographic controls added in the baseline. Acronyms read as follows: FDI is *Fratelli d'Italia*, RW and LW are respectively "right-wing" and "left-wing". PD, M5S and PDL, FI stand for *Democratic Party*, *Movimento 5 Stelle*, and *Popolo della Libertà*, *Forza Italia*, respectively. Stars correspond to the following p-values: * p < .10, ** p < .05, *** p < .001.

6. Mechanisms and Robustness

6.1. Mechanisms

To further investigate these dynamics, I explore whether the demand of readers for ideology had any role in explaining the findings observed. In particular, I introduce two outcome variables: 1) a measure of Google Trends traffic towards the local outlets in sampled municipalities; 2) a measure of local news penetration, captured as relative traffic with respect to National news. The first variable is a weighted sum of monthly variation in outlet traffic, with weights varying by the importance of the news source in the municipality in each time period.

$$\text{Traffic}_{mt} = \sum_{s \in S} w_{smt} \text{traffic}_{st} \quad (6)$$

The second variable is a weighted sum of monthly variation in outlet traffic, relative to traffic for a mainstream National newspaper, with weights varying by the importance of the news source in the municipality in each

time period.

$$\text{Penetration}_{mt} = \sum_{s \in S} w_{smt} \text{traffic}_{st} / \text{traffic}_{S_{NAT}t} \quad (7)$$

In the formulas, traffic_{st} is the rating assigned by Google Trends to the keyword representing outlet s , w_{smt} is the same weight used for the slant measure; traffic_{St} is the rating assigned by Google Trends to the keyword representing mainstream national newspaper S_{NAT} : *La Stampa* as a benchmark and *La Repubblica*, *Il Fatto Quotidiano*, *Il Giornale* as other outlets considered with results in the Appendix.²⁷ A visualization of the measures created appears in the Appendix Figure A.8. Note that the last column of Table 2 explores traffic as an outcome in the baseline diff-in-diff specification. Post-border closures, traffic decreases with proximity to the border, thus taking the same direction of anti-immigrant slant.²⁸

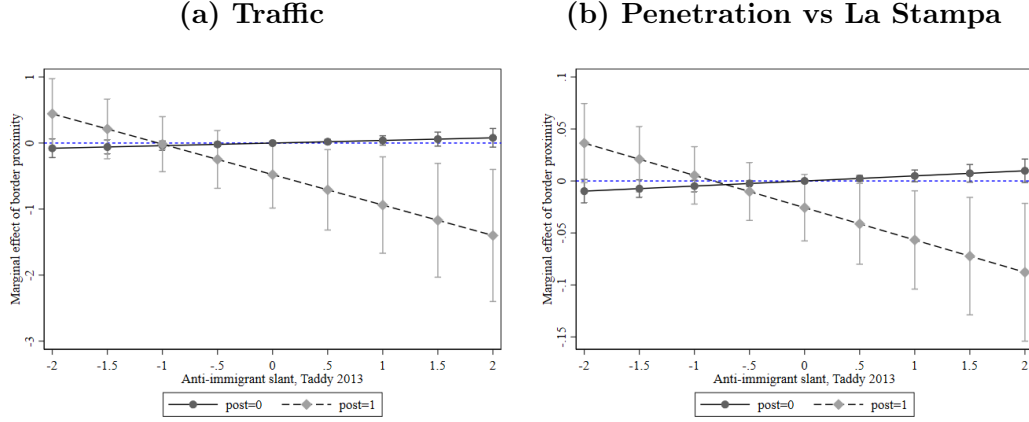
Additionally, considering both traffic and penetration as dependent variables, I regress each of them on a triple interaction between border proximity, a post-treatment dummy and slant. Figure 8 plots the results of this specification. The figure shows the marginal effects of border proximity on traffic (plot a) and penetration (plot b) on the y-axis. On the x-axis, the figure shows different levels of anti-immigrant slant. The plotted lines and markers display how marginal effects vary with slant being more or less anti-immigrant, in the pre (darker continuous line) and post-event (lighter scattered line) periods. Focusing on plot a, we can observe that prior to the border closures, the marginal effects of border proximity on traffic are close to zero and non-significant. This stays uniform at all levels of anti-immigrant slant. After the border policy, proximity to the border has a negative effect on traffic, if slant is more anti-immigrant. Specifically, the marginal effects of border proximity turn significantly negative as long as the anti-immigrant slant is not extremely low and increase in magnitude where the slant is more anti-immigrant. A similar pattern appears when looking at penetration as an outcome, in plot b. In other words, far from the border, higher levels of anti-immigrant slant (observed in the baseline results) appear to increment user traffic towards local news.

The link between slant, distance to the border and local news traffic and penetration could be justified by local news targeting their content for the local readerships. Further from the border, readers do not experience events directly. They form (or hold) beliefs that differ from those of readers at the border and look for their views to be matched in the news. Local news react by adjusting their offer to satisfy this demand. By adjusting the slant, they appear to gain market from national news, which is less able to accommodate specific areas. Content adjustments by local news based on reader targeting and national news coverage are also documented in George and Waldfogel (2006), for the US.

²⁷See Askitas (2015) for a review on the scope of Google Trends data for detecting relevant socio-economic trends with high time-frequency, worldwide.

²⁸Appendix Figure A.9 complements this finding, by showing results from an event study specification, where traffic and penetration are the outcomes of interest.

Figure 8. Effects on traffic and penetration by levels of slant



Notes: marginal effects from a triple-differences specification. Local news traffic (left) and penetration with respect to national outlet La Stampa (right) are the outcomes of interest. Figure A.11 in the Appendix shows additional results for other National newspapers; namely: La Repubblica, Il Fatto Quotidiano, Il Giornale.

News at the border is subject to an audience that also directly observes the events.²⁹ Close to the events, slant is less anti-immigrant. This finding may indicate an exclusively favorable response from natives (then reflected in the news), or alternatively a mixed ideological response. Voting patterns seem to point towards this second alternative. Along such interpretation, readers' heterogeneous beliefs can lower anti-immigrant bias in newspapers, if, as in the model of Mullainathan and Shleifer (2005), readers seek confirmatory news, that reinforces their beliefs, and news outlets accommodate them. In the authors' framework, reader diversity is a crucial condition to reach accuracy of information overall, because biased news accommodates readers in both ideological directions. In other words, the presence of mixed beliefs of the readers leads to news whose bias directions neutralize in aggregate. Conversely, away from the policy, beliefs are not as bounded by the exposure to real events, which are distant and thus less observed. By accommodating a uniform ideological bias, local news succeeds in penetrating the market.

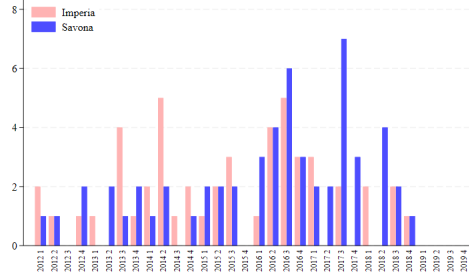
In accordance with this interpretation, Steinmayr (2021) discusses the opposite implications of contact versus mere exposure of natives to migrants.³⁰ Consistently with the contact-theory framework, natives would have responded on average more favorably closer to the border. Anecdotally, several NGOs were present at the border in the aftermath of the borders' closure. This could have been a further stimulus for the existence

²⁹Gentzkow et al. (2018) consider a model of trust in information sources where individuals both receive information from news and from direct exposure to an event and decide on the information accuracy depending on their own ideology. This allows them to trust more news contents that validate their views.

³⁰This framework is based on the work on social psychology by Allport (1954).

of some positive responses to migration in the area. Figure 9 is in line with the contact-theory mechanism. It reports a count measure of hate crimes in the two provinces of interest. This figure builds on Romarri (2020) who investigates the impact of far-right mayors on hate crimes, employing a database collected by NGO Lunaria.³¹ Using the same data, Figure 9 depicts the time variation of hate crimes, aggregated by province. An emerging pattern reflects that, past 2015, hate crime prevalence grew relatively more in the province of Savona, which is further from the border.³²

Figure 9. Hate-crimes over time, by province



Notes: Time series of hate crimes, compiled by database Lunaria, used in Romarri (2020).

6.2. Sensitivity to identification assumptions and other robustness

6.2.1. Identification and Placebos

The analysis presented so far accounts for several violations of the parallel trends. All specifications from the last column in the baseline onward include province-specific time effects. Estimates also absorb unobserved heterogeneity that is captured by possibly confounding geographic characteristics; namely latitude, longitude, and altitude, interacting with the post-events dummy. The addition of several covariates after studying the balance of the sample is also a measure to improve comparability. Event-study graphs with this specification do not point to significant pre-policy effects. As a robustness to the fixed effect baseline specification, Table B.4 adds time trends at the municipality level to the preferred specification and displays coefficients in line with the main findings. A concern, however, relates to the presence of a continuous treatment in the context of a diff-in-diff methodology. Callaway et al. (2021) observe that when the treatment is continuous, estimating the average causal response (ACR) of a treatment dose d by two-way fixed effects (TWFE) presents some limitations. As a side note, this analysis is spared from further complications that also arise when treatment timing is staggered. Here instead, the border closure occurred in one specific period, impacting the whole sample at the same time. One limitation in the present analysis involves the weights that TWFE estimates

³¹www.cronachediordinariorazzismo.org

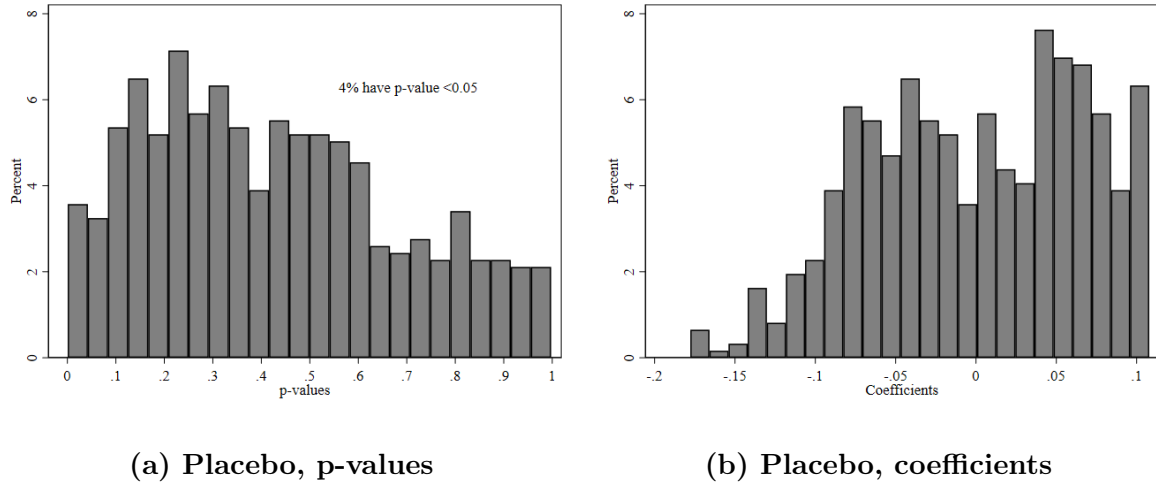
³²Note the count variable in the period of interest has low variation. This lack of heterogeneity motivates the choice of not running a regression for this outcome.

assign to each dose. If the goal is to obtain a weighted average of the ACR, with weights matching the real dose distribution, TWFE estimates tend to give too much importance to mean-dose observations, thus under-representing the distribution’s tails. To address this issue, in a battery of sensitivity checks, I look at several alternative functional forms of distance, including binary distance specifications. As these different measurements change how the distance (dose) distribution looks like, one can observe how sensitive are TWFE parameters to the use of inverse distance.³³

An additional issue of TWFE estimates relates to the parallel trends assumption, which is required to hold at every treatment-dose comparison. If this stronger assumption doesn’t hold, TWFE estimates would include a selection bias term arising from the different paths that different doses would have followed in a counterfactual state of nature. The policy of border push-backs was introduced by the French authorities, which respond to different governments and jurisdictions than the one governing the sampled areas. Distance to the border would hardly reflect a selection into treatment for natives in the Italian areas. The use of distance induces a reduced-form estimate of the direct exposure to migrants, without the element of selection into exposure. However, distance to the border may also proxy several unobserved factors that may harm the full comparability of municipalities, beyond those that have been included in the controls, and even if the provinces in the sample were targeted specifically to ensure that observations were more homogeneous along several factors. To challenge the validity of using proximity to the border as an appropriate proxy for treatment intensity, a placebo analysis is performed as follows. In the baseline specification, the actual inverse distance to the border town of Ventimiglia is replaced by the inverse distance to a set of Italian towns taken randomly (outside the North West region of Italy, to avoid an excessive collinearity between distance measures). Geographic coordinates for Italian towns, villages and cities derive from OpenStreetMap (2023). A random 10% sample of the collection of Italian towns in Open Street Map is retained for the placebo. Commuting times are calculated with the R package *osrmTable* (Mark Padgham et al., 2017). A placebo regression is repeated for every extracted town. The histogram of Figure 10 plots the distribution of resulting p-values on the left. On the right of the figure, the coefficients’ values are plotted. A 4% of the random distances leads to a significant estimate, indicating that baseline estimates do not over-reject the null hypothesis for the coefficient of interest. The continuous treatment adopted in the baseline therefore appears to well-identify the analyzed impact.

³³In the Appendix, Table B.5 shows robustness tests with respect to alternative functional forms (linear, logarithmic, quadratic) for commuting distance. Figure A.12 shows coefficients from binary treatment diff-in-diff regressions, in which a binary treatment equals 1 at different distance thresholds reported in the x-axis.

Figure 10. Placebo coefficients



6.2.2. Other Robustness Tests

An additional battery of robustness tests is performed to allow the possible exclusion of a municipality (i.e., a panel unit). Because the set of municipalities under study is restricted to a small sample, a test is conducted to make sure that the results on the three outcomes are not driven by the presence of a single municipality. Figures A.13, and A.14 reproduce both betas and associated t-statistics for all possible permutations of such exclusions. The figures indicate that both the sign and significance of the effects are maintained in all these alternatives. On a similar ground, Figure A.15 summarizes results when a month is excluded from the estimation. Results persist in sign and significance to each of these possible alternatives. A related concern involves the use of an imputation method to extend the original geographic distribution of news, retrieved by Google Trends. Note without imputing the sample substantially shrinks to 38 municipalities. Table B.6 shows that even when restricting the sample to the original non-imputed data, the significance and direction of results are maintained. Reassuringly, slant coefficients are very much in line with the baseline estimates.

Finally, I run an additional placebo test for the validity of the algorithm predicting anti-immigrant slant. I build a training sample as in the original training sample, but I shuffle the 0-1 labels randomly. Then, I make slant predictions on the articles in local news as done for the original measure. I repeat the process and create 1000 placebo slant variables.³⁴ Figure A.16 plots the distribution of resulting p-values on the left. On the right of the figure, the coefficients' values are plotted. The picture shows insignificant results for the placebo-slant measure. This confirms that it is the original training sample, rather than a random label assignment, that leads to significant results in the baseline.

³⁴This procedure leads to a vast number of prediction probabilities to be exactly 1/2: these entries are randomly assigned a 0 or a 1.

7. Conclusion

In this paper, I examine how local news markets can be shaped by salient events in the context of the recent European migration crisis and I document a relationship between news-market outcomes and trends in the local political landscape. Specifically, I provide new evidence that local shocks in migrant presence impact discourse in local news, both in terms of quantity (*agenda-setting*) and quality (*framing*), though in opposite directions. The causal identification of these effects follows from a specific geographic and historical setting. A local shock in the presence of migrants at the French-Italian border in June 2015 emerged from the establishment of a militarized push-back system by the French authorities. The coastal border had previously constituted an important gateway for precarious migrants, intending to leave Italy for other European destinations.

Understanding how news speech evolves is key to unraveling aspects related to the evolution of citizens' attitudes (Djourelouva, 2023; Keita et al., 2023) and language analysis can retrieve rich information on the political landscape beyond what the observed electoral outcomes can capture (Gentzkow et al., 2019b). For example, Campante et al. (2018) document how the spread of the internet in Italy influenced dissatisfied voters. Their behavior changed over time and only ultimately resulted in voting for the anti-elite Five-Star Movement.

I rely on a rich data collection of articles to investigate the outputs of news outlets. Evidence suggests the border push-back events translated into a salient shock in the media: municipalities closer to the event received more migrant-related news, and coverage decays rapidly with distance.

The share of anti-immigrant news, however, got relatively higher in the municipalities further away from the border, while it lowered at the border. I propose some rationalizations for this finding. I find an important role for demand for local news and its market penetration, as measured by the online traffic for local news and its relation with traffic towards national outlets. The relative interest in local news turns out to change in the same direction as slant, after the border policy. Additionally, in the least exposed areas, where the slant is more anti-immigrant, local news gains more market. If local news readers close to the border become more heterogeneous in terms of their views on migration, local demand for anti-immigrant contents may be counterbalanced, at the border, by the presence of favourable responses. This mechanism can relate to the contact versus exposure paradigm. Lower slant levels in the areas closest to the border would reflect the presence of some empathetic attitudes as more contact occurred with the migrants' precarious situation, whilst municipalities further away, lacking this contact component, responded more uniformly negatively. This line of interpretation is matched by: i) overall trends in hate crime events, measured à la Romarri (2020) and ii) vote shares differentials for the general elections, where higher heterogeneity in the support appears in the vicinity of the border. Baseline results prove robust to several sensitivity checks on the specification of the model and the measurement of the main variables.

Despite its local setting, the policy relevance of this contribution is broader. This study relates to the importance of international coordination in migration policies (Cimade, 2018; Giordani and Ruta, 2013):

as an example of cross-country tensions in the political debate on internal borders policies, a provocative tweet of anti-immigrant party FDI leader Giorgia Meloni stated: “Important notice to migrants! France, Germany, and Luxembourg have declared that closing ports violates international law and that rescue cannot be criminalized. Head to their *borders*, you will be received with open arms by these welcoming states.”³⁵ Finally, this paper focuses on local news and employs standard text-analysis *bag-of-words* methods in the prediction of a directed measure of slant. An important area for future research will involve the exploration of other relevant media and the use of promising, evolving techniques to possibly capture even deeper mappings between text nuances and ideology.

³⁵Avviso importante ai migranti! Francia, Germania e Lussemburgo hanno dichiarato che chiudere i porti viola il diritto internazionale e che il soccorso non può essere criminalizzato. Dirigetevi verso le loro frontiere, sarete accolti a braccia aperte da questi Stati accoglienti. Source:<https://twitter.com/giorgiameloni/status/1145015881195540483>, posted in June 2019, retrieved in October 2021.

During the preparation of this work, the author used OpenAI’s chatGPT to improve the readability of some sentences. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Bibliography

- Aksoy, Cevat Giray and Panu Poutvaara (2021) “Refugees’ and irregular migrants’ self-selection into Europe,” *Journal of Development Economics*, 152, 102681, <https://doi.org/10.1016/j.jdevco.2021.102681>.
- Alesina, Alberto and Marco Tabellini (2024) “The Political Effects of Immigration: Culture or Economics?” *Journal of Economic Literature*, 62 (1), 5–46, 10.1257/jel.20221643.
- Allport, Gordon W (1954) “The nature of prejudice,” *Reading: Addison-Wesley*, 129.
- Altındağ, Onur and Neeraj Kaushal (2020) “Do refugees impact voting behavior in the host country? Evidence from Syrian refugee inflows to Turkey,” *Public Choice*, 1–30, Publisher: Springer.
- Anderson, Simon P and John McLaren (2012) “Media Mergers and Media Bias with Rational Consumers,” *Journal of the European Economic Association*, 10 (4), 831–859, Publisher: Oxford University Press.
- Askitas, Nikolaos (2015) “Google search activity data and breaking trends,” *IZA World of Labor* (206), Publisher: Bonn: Institute for the Study of Labor (IZA).
- Barone, Guglielmo, Alessio D’Ignazio, Guido de Blasio, and Paolo Naticchioni (2016) “Mr. Rossi, Mr. Hu and politics. The role of immigration in shaping natives’ voting behavior,” *Journal of Public Economics*, 136, 1–13, Publisher: Elsevier.
- Battiston, Giacomo (2022) “Rescue on Stage: Border Enforcement and Public Attention in the Mediterranean Sea,” "Marco Fanno" Working Papers 0292, Dipartimento di Scienze Economiche "Marco Fanno", <https://ideas.repec.org/p/pad/wpaper/0292.html>.
- Benesch, Christine, Simon Loretz, David Stadelmann, and Tobias Thomas (2019) “Media coverage and immigration worries: Econometric evidence,” *Journal of Economic Behavior & Organization*, 160, 52–67, Publisher: Elsevier.
- Blei, David M, Andrew Y Ng, and Michael I Jordan (2003) “Latent Dirichlet Allocation,” *Journal of Machine Learning Research*, 3 (Jan), 993–1022.
- Bouchet-Valat, Milan (2020) “Package ‘SnowballC’,” *R package version 0.7. 0*.
- Branton, Regina P and Johanna Dunaway (2009) “Spatial Proximity to the U.S.-Mexico Border and Newspaper Coverage of Immigration Issues,” *Political Research Quarterly*, 62 (2), 289–302, Place: Los Angeles, CA Publisher: Sage Publications.

- Bratti, Massimiliano, Claudio Deiana, Enkelejda Havari, Gianluca Mazzarella, and Elena Claudia Meroni (2020) “Geographical proximity to refugee reception centres and voting,” *Journal of Urban Economics*, 120, Publisher: Elsevier Inc.
- Bursztyn, Leonardo, Thomas Chaney, Tarek A. Hassan, and Aakaash Rao (2024) “The Immigrant Next Door,” *American Economic Review*, 114 (2), 348–384, 10.1257/aer.20220376.
- Callaway, Brantly, Andrew Goodman-Bacon, and Pedro HC Sant’Anna (2021) “Difference-in-Differences with a Continuous Treatment,” *arXiv preprint arXiv:2107.02637*.
- Campante, Filipe, Ruben Durante, and Francesco Sobbrío (2018) “Politics 2.0: The Multifaceted Effect of Broadband Internet on Political Participation,” *Journal of the European Economic Association*, 16 (4), 1094–1136, Publisher: Oxford University Press.
- Campo, Francesco, Sara Giunti, and Mariapia Mendola (2021) “The Refugee Crisis and Right-Wing Populism: Evidence from the Italian Dispersal Policy,” *IZA DP No. 14084*.
- Capitani, Guilia (2018) “Nowhere But Out: The failure of France and Italy to help refugees and other migrants stranded at the border in Ventimiglia,” 10.21201/2018.2708, Publication Title: Oxfam Policy & Practice.
- Chen, Song Xi (2000) “Probability Density Function Estimation Using Gamma Kernels,” *Annals of the Institute of Statistical Mathematics*, 52 (3), 471–480, Publisher: Kluwer Academic Publishers.
- Chouliaraki, Lilie, Myria Georgiou, Rafal Zaborowski, and WA Oomen (2017) “The European ‘migration crisis’ and the media: a cross-European press content analysis.”
- Christ, Oliver, Katharina Schmid, Simon Lolliot et al. (2014) “Contextual effect of positive intergroup contact on outgroup prejudice,” *Proceedings of the National Academy of Sciences*, 111 (11), 3996–4000, 10.1073/pnas.1320901111, Publisher: Proceedings of the National Academy of Sciences.
- Cimade, L (2018) “Dedans, Dehors: Une Europe Qui S’enferme,” *Report, Cimade, France, June*.
- Colombeau, Sara Casella (2020) “Crisis of Schengen? The effect of two ‘migrant crises’ (2011 and 2015) on the free movement of people at an internal Schengen border,” *Journal of Ethnic and Migration Studies*, 46 (11), 2258–2274, 10.1080/1369183X.2019.1596787, Publisher: Routledge _eprint: <https://doi.org/10.1080/1369183X.2019.1596787>.
- de Coulon, Augustin, Dragos Radu, and Max Friedrich Steinhardt (2016) “Pane e Cioccolata: The Impact of Native Attitudes on Return Migration,” *Review of International Economics*, 24 (2), 253–281, Publisher: Wiley Online Library.

- DellaVigna, Stefano, Ruben Enikolopov, Vera Mironova, Maria Petrova, and Ekaterina Zhuravskaya (2014) “Cross-Border Media and Nationalism: Evidence from Serbian Radio in Croatia,” *American Economic Journal: Applied Economics*, 6 (3), 103–32.
- DellaVigna, Stefano and Ethan Kaplan (2007) “The Fox News Effect: Media Bias and Voting,” *The Quarterly Journal of Economics*, 122 (3), 1187–1234, Publisher: MIT Press.
- Dinas, Elias, Konstantinos Matakos, Dimitrios Xeferis, and Dominik Hangartner (2019) “Waking Up the Golden Dawn: Does Exposure to the Refugee Crisis Increase Support for Extreme-Right Parties?” *Political Analysis*, 27, 1–11, 10.1017/pan.2018.48.
- Djourelouva, Milena (2023) “Persuasion through Slanted Language: Evidence from the Media Coverage of Immigration,” *American Economic Review*, 113 (3), 800–835, 10.1257/aer.20211537.
- Durante, Ruben, Paolo Pinotti, and Andrea Tesei (2019) “The Political Legacy of Entertainment TV,” *American Economic Review*, 109 (7), 2497–2530.
- Dustmann, Christian, Francesco Fasani, Tommaso Frattini, Luigi Minale, and Uta Schönberg (2017) “On the economics and politics of refugee migration,” *Economic Policy*, 32 (91), 497–550, Publisher: Oxford University Press.
- Dustmann, Christian, Kristine Vasiljeva, and Anna Piil Damm (2019) “Refugee migration and electoral outcomes,” *The Review of Economic Studies*, 86 (5), 2035–2091, Publisher: Oxford University Press.
- D’Alimonte, Roberto (2013) “The Italian elections of February 2013: the end of the Second Republic?” *Contemporary Italian Politics*, 5 (2), 113–129, Publisher: Taylor & Francis.
- Edo, Anthony, Yvonne Giesing, Jonathan Öztunc, and Panu Poutvaara (2019) “Immigration and electoral support for the far-left and the far-right,” *European Economic Review*, 115, 99–143, Publisher: Elsevier.
- Facchini, Giovanni, Anna Maria Mayda, and Riccardo Puglisi (2017) “Illegal immigration and media exposure: evidence on individual attitudes,” *IZA Journal of Development and Migration*, 7 (1), 1–36, Publisher: SpringerOpen.
- Falck, Oliver, Robert Gold, and Stephan Heblich (2014) “E-lections: Voting Behavior and the Internet,” *American Economic Review*, 104 (7), 2238–65.
- Flaxman, Seth, Sharad Goel, and Justin M Rao (2016) “Filter Bubbles, Echo Chambers, and Online News Consumption,” *Public Opinion Quarterly*, 80 (S1), 298–320, Publisher: Oxford University Press.
- Friedman, Jerome, Trevor Hastie, and Rob Tibshirani (2010) “Regularization Paths for Generalized Linear Models via Coordinate Descent,” *Journal of Statistical Software*, 33 (1), 1, Publisher: NIH Public Access.

- Friedman, Jerome, Trevor Hastie, Rob Tibshirani, and others (2009) “glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models,” *R package version*, 1 (4), 1–24.
- Gallegos Torres, Katia (2023) “The 2015 refugee inflow and concerns over immigration,” *European Journal of Political Economy*, 78, 102323, 10.1016/j.ejpoleco.2022.102323.
- Gamalerio, Matteo and Margherita Negri (2023) “Not welcome anymore: the effect of electoral incentives on the reception of refugees,” *Journal of Economic Geography*, 23 (4), 901–920, 10.1093/jeg/lbad002, _eprint: <https://academic.oup.com/joeg/article-pdf/23/4/901/50859585/lbad002.pdf>.
- Gentzkow, Matthew (2006) “Television and Voter Turnout,” *The Quarterly Journal of Economics*, 121 (3), 931–972, Publisher: Oxford University Press.
- (2009) “Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions,” *American Economic Journal: Applied Economics*, 1 (2), 35–52.
- Gentzkow, Matthew, Bryan Kelly, and Matt Taddy (2019a) “Text as data,” *Journal of Economic Literature*, 57 (3), 535–74.
- Gentzkow, Matthew and Jesse M Shapiro (2006) “Media bias and reputation,” *Journal of Political Economy*, 114 (2), 280–316, Publisher: The University of Chicago Press.
- Gentzkow, Matthew and Jesse M. Shapiro (2010) “What Drives Media Slant? Evidence From U.S. Daily Newspapers,” *Econometrica*, 78 (1), 35–71, Place: Oxford, UK Publisher: Blackwell Publishing Ltd.
- Gentzkow, Matthew, Jesse M. Shapiro, and Daniel F. Stone (2015) “Chapter 14 - Media Bias in the Marketplace: Theory,” in Anderson, Simon P., Joel Waldfogel, and David Strömberg eds. *Handbook of Media Economics*, 1 of Handbook of Media Economics, 623–645: North-Holland, <https://doi.org/10.1016/B978-0-444-63685-0.00014-0>, ISSN: 2213-6630.
- Gentzkow, Matthew, Jesse M Shapiro, and Matt Taddy (2019b) “Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech,” *Econometrica*, 87 (4), 1307–1340, Publisher: Wiley Online Library.
- Gentzkow, Matthew, Michael B Wong, and Allen T Zhang (2018) “Ideological Bias and Trust in Information Sources,” *Unpublished Manuscript*.
- George, Lisa M. and Joel Waldfogel (2006) “The “New York Times” and the Market for Local Newspapers,” *The American Economic Review*, 96 (1), 435–447, <http://www.jstor.org/stable/30034376>, Publisher: American Economic Association.

- Gessler, Theresa, Gerg\Ho Tóth, and Johannes Wachs (2021) “No country for asylum seekers? How short-term exposure to refugees influences attitudes and voting behavior in Hungary,” *Political Behavior*, 1–29, Publisher: Springer.
- Giordani, Paolo E and Michele Ruta (2013) “Coordination failures in immigration policy,” *Journal of International Economics*, 89 (1), 55–67, Publisher: Elsevier.
- Groseclose, Tim and Jeffrey Milyo (2005) “A measure of media bias,” *The Quarterly Journal of Economics*, 120 (4), 1191–1237, Publisher: MIT Press.
- Halla, Martin, Alexander F Wagner, and Josef Zweimüller (2017) “Immigration and voting for the far right,” *Journal of the European Economic Association*, 15 (6), 1341–1385, Publisher: Oxford University Press.
- Hangartner, Dominik, Elias Dinas, Moritz Marbach, Konstantinos Matakos, and Dimitrios Xefteris (2019) “Does Exposure to the Refugee Crisis Make Natives More Hostile?” *American Political Science Review*, 113 (2), 442–455, Publisher: Cambridge University Press.
- Hornik, Kurt and Bettina Grün (2011) “topicmodels: An R package for fitting topic models,” *Journal of Statistical Software*, 40 (13), 1–30, Publisher: American Statistical Association.
- Iannelli, Laura, Bianca Biagi, and Marta Meleddu (2021) “Public opinion polarization on immigration in Italy: the role of traditional and digital news media practices,” *The Communication Review*, 24 (3), 244–274, 10.1080/10714421.2021.1972720, Publisher: Routledge.
- Keita, Sekou, Thomas Renault, and Jérôme Valette (2023) “The Usual Suspects: Offender Origin, Media Reporting and Natives’ Attitudes Towards Immigration,” *The Economic Journal*, uead059, 10.1093/ej/uead059, _eprint: <https://academic.oup.com/ej/advance-article-pdf/doi/10.1093/ej/uead059/51797584/uead059.pdf>.
- Lonsky, Jakub (2021) “Does immigration decrease far-right popularity? Evidence from Finnish municipalities,” *Journal of Population Economics*, 34 (1), 97–139, Publisher: Springer.
- Mark Padgham, Bob Rudis, Robin Lovelace, and Maëlle Salmon (2017) “osmdata,” *Journal of Open Source Software*, 2 (14), 305, 10.21105/joss.00305, Publisher: The Open Journal.
- Mayda, Anna Maria, Giovanni Peri, and Walter Steingress (2016) “Immigration to the US: A Problem for the Republicans or the Democrats?” Technical report, National Bureau of Economic Research.
- Mousa, Salma (2020) “Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq,” *Science*, 369 (6505), 866–870, 10.1126/science.abb3153, Publisher: American Association for the Advancement of Science.
- Mullainathan, Sendhil and Andrei Shleifer (2005) “The Market for News,” *American Economic Review*, 95 (4), 1031–1053.

- OpenStreetMap (2023) "Italy's Sub Regions retrieved from <https://download.geofabrik.de/europe/italy.html> on 2023-05-03," Published: <https://www.openstreetmap.org>.
- Otto, Alkis Henri and Max Friedrich Steinhardt (2014) "Immigration and election outcomes-Evidence from city districts in Hamburg," *Regional Science and Urban Economics*, 45, 67–79, Publisher: Elsevier.
- Peracchi, Silvia (2025) "Replication data for: "Migration Crisis in the Local News: Evidence from the French-Italian Border."," 10.5281/zenodo.16738591, [Data set]. In *Journal of Urban Economics* (Version V1).
- Pettigrew, Thomas (1998) "Intergroup Contact Theory," *Annual Review of Psychology*, 49, 65–85, 10.1146/annurev.psych.49.1.65.
- Prat, Andrea and David Strömberg (2013) "The Political Economy of Mass Media," *Advances in Economics and Econometrics: Theory and Applications, Proceedings of the Tenth World Congress of the Econometric Society*.
- Romarri, Alessio (2020) "Do Far-Right Mayors Increase the Probability of Hate Crimes? Evidence From Italy," December, 10.2139/ssrn.3506811, Place: Rochester, NY Type: SSRN Scholarly Paper.
- Scheufele, Dietram A and David Tewksbury (2007) "Framing, Agenda Setting, and Priming: The Evolution of Three Media Effects Models," *Journal of Communication*, 57 (1), 9–20, Publisher: Oxford University Press.
- Schindler, David and Mark Westcott (2021) "Shocking Racial Attitudes: Black G.I.s in Europe," *The Review of Economic Studies*, 88 (1), 489–520, 10.1093/restud/rdaa039.
- Snyder Jr, James M and David Strömberg (2010) "Press Coverage and Political Accountability," *Journal of Political Economy*, 118 (2), 355–408, Publisher: The University of Chicago Press.
- Steinmayr, Andreas (2021) "Contact versus Exposure: Refugee Presence and Voting for the Far Right," *The Review of Economics and Statistics*, 103 (2), 310–327, <https://ideas.repec.org/a/tpr/restat/v103y2021i2p310-327.html>.
- Strömberg, David (2004) "Radio's Impact on Public Spending," *The Quarterly Journal of Economics*, 119 (1), 189–221, Publisher: MIT Press.
- Taddy, Matt (2013) "Multinomial inverse regression for text analysis," *Journal of the American Statistical Association*, 108 (503), 755–770, Publisher: Taylor & Francis.
- Tebaldi, Mauro (2014) "Il potere di esternazione del capo dello Stato nella Seconda Repubblica," *Comunicazione Politica*, 14 (3), 485–506, Publisher: Società editrice il Mulino.

Figure A.2. Meal counts distributed by Caritas in 2017



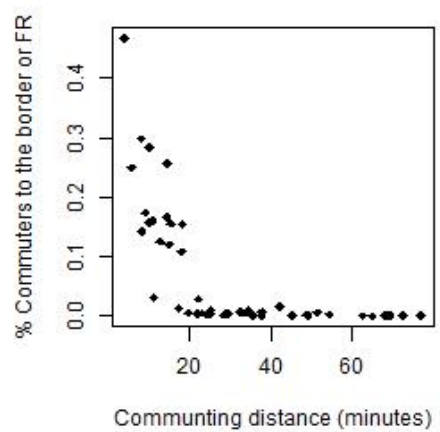
Notes: Stylized statistics on the presence of asylum seekers and migrants that were offered a meal by local NGO Caritas Intemelja. Credits for this graph go to Caritas Intemelja.

Figure A.3. Comparison cloud. Terms with highest vs. lowest right-wing affinity



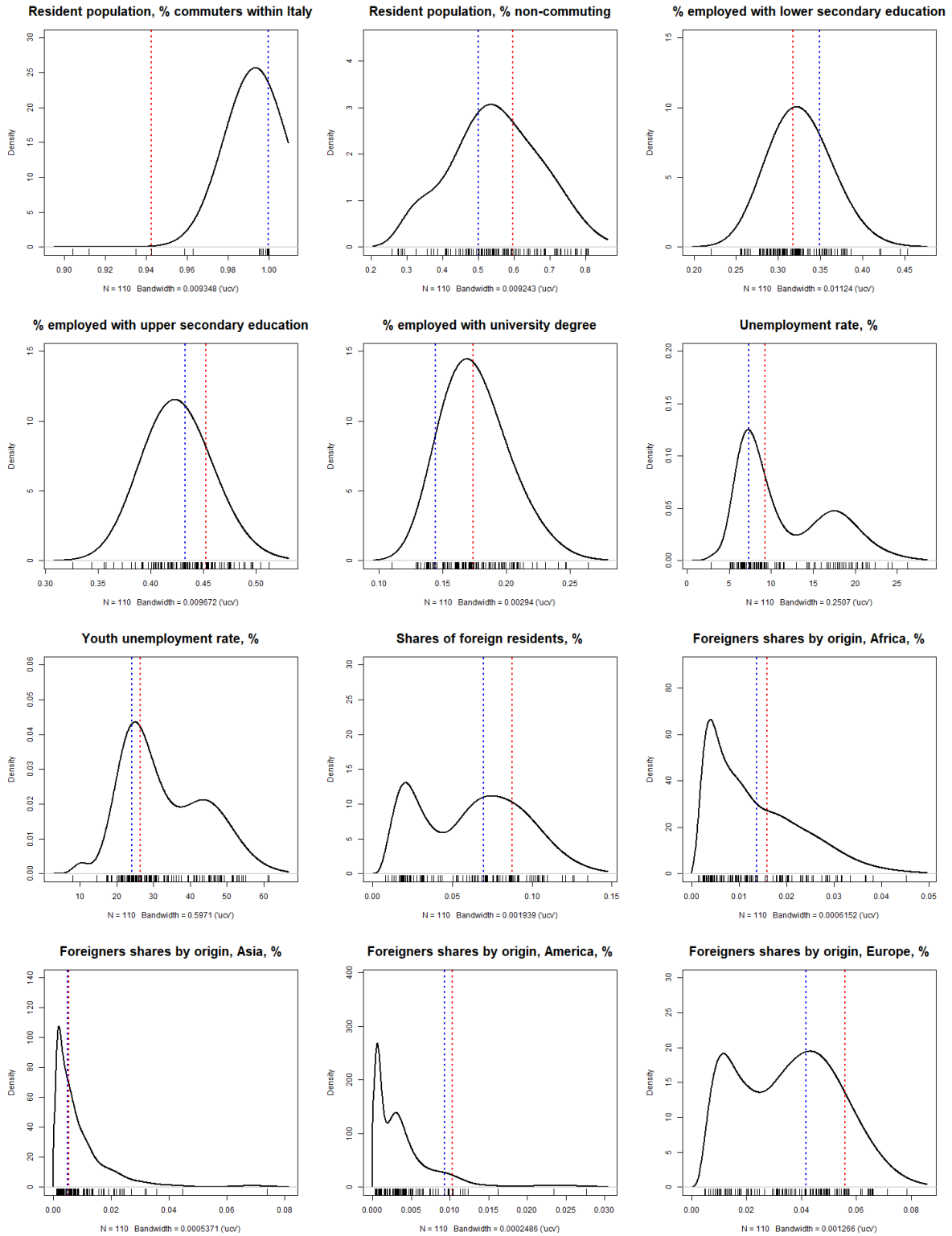
Notes: Exploration of the predictions resulting from the baseline classification approach. Prediction probabilities are employed to subset the sampled articles into least likely (top 10% of the assigned probability, at the top, in red) and most likely (bottom 10% of the assigned probability, at the bottom, in blue) anti-immigrant.

Figure A.4. Distance to the border and commuting to and across the border



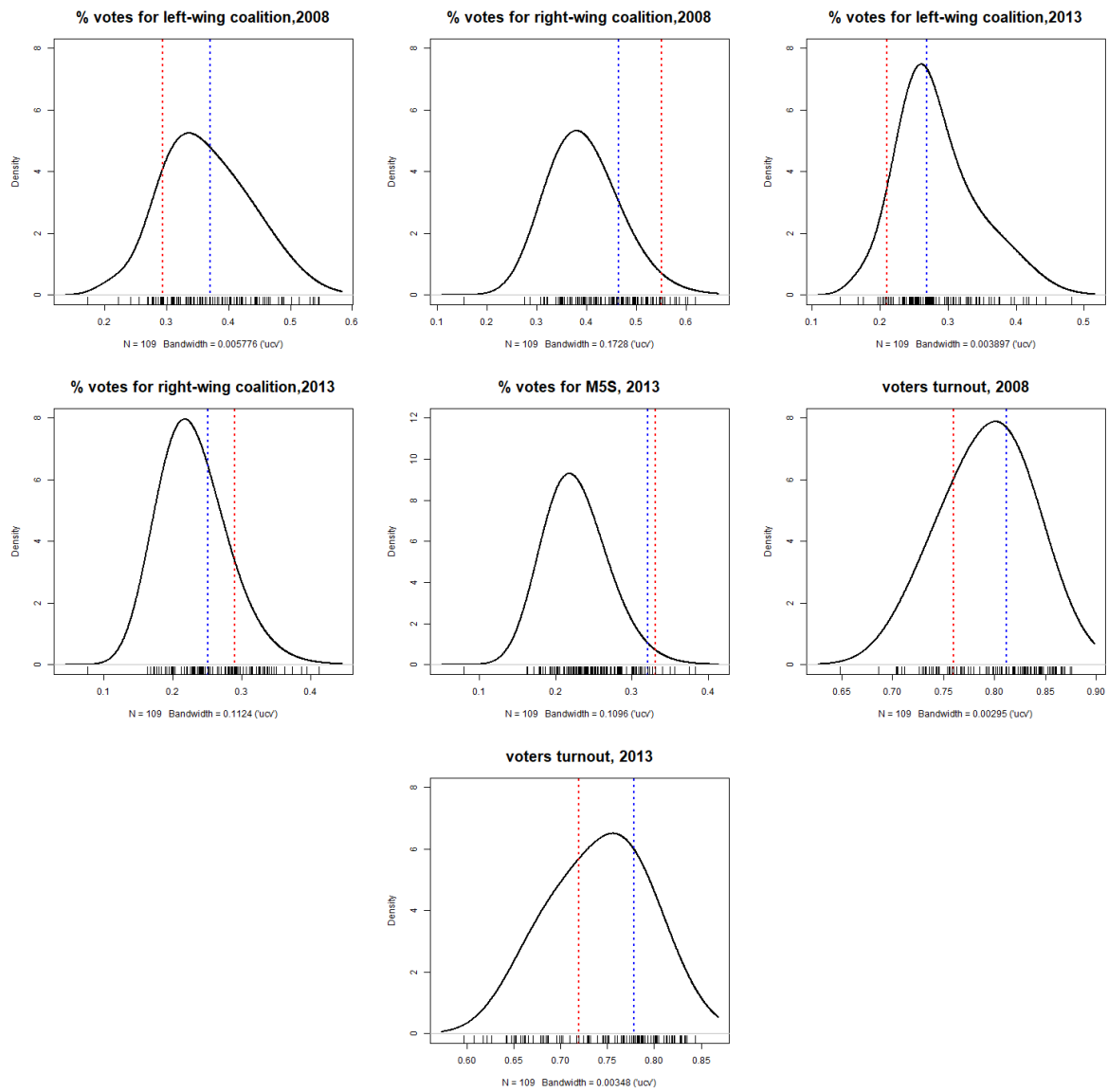
Notes: Source: Commuters' matrices (*Matrici del pendolarismo*), ISTAT, based on the 2011 census' data.

Figure A.5. Provinces of interest versus rest of Italy, 2011



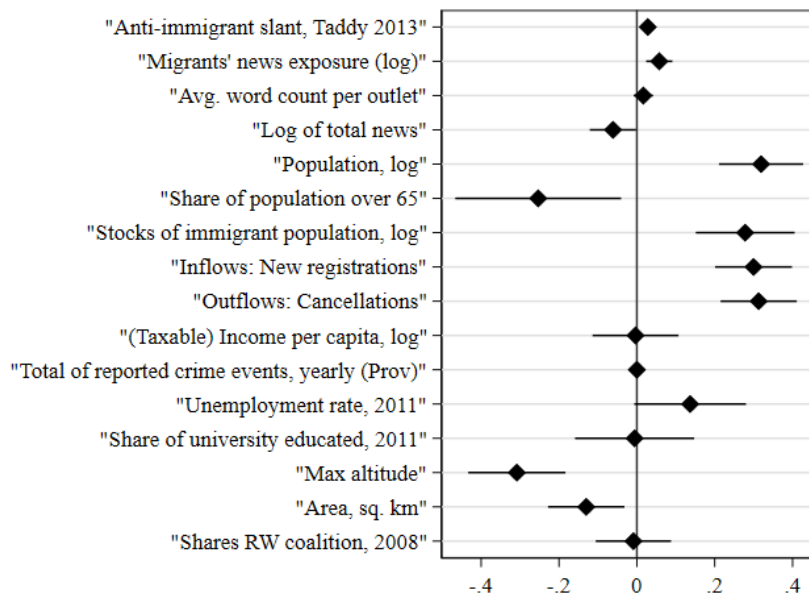
Notes: These graphs show where Imperia (red dotted line) and Savona (blue dotted line) lie within the distribution of Italian provinces on a set of socio-demographic characteristics. Data are sourced from ISTAT, and belong to the Census 2011 dataset. Density graphs are recovered from a univariate kernel density estimation that uses gamma kernel functions as in Chen (2000), where such kernel choice accounts for the positive range of the data. Bandwidth selection is based on unbiased cross-validation (UCV).

Figure A.6. Provinces of interest versus rest of Italy, continued



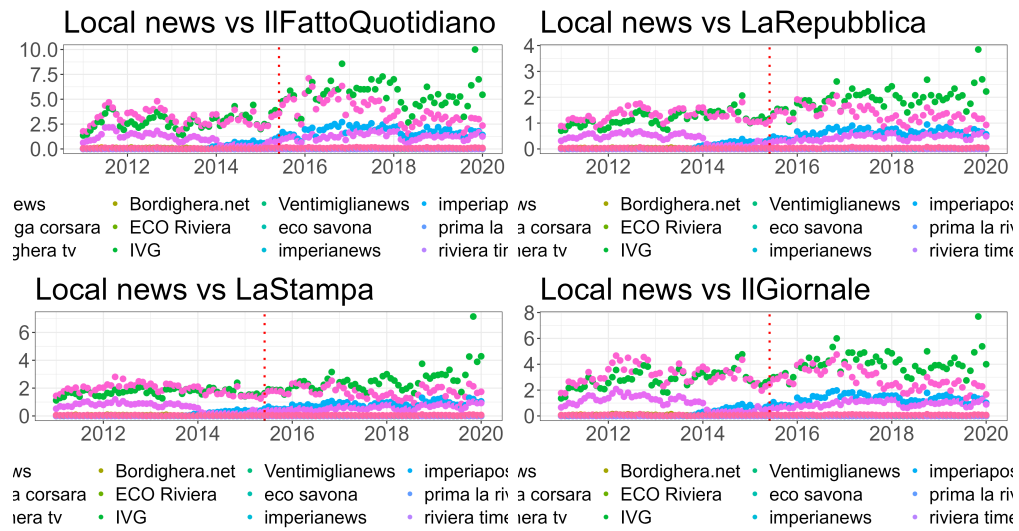
Notes: Continuation of Figure A.5.

Figure A.7. Balance check



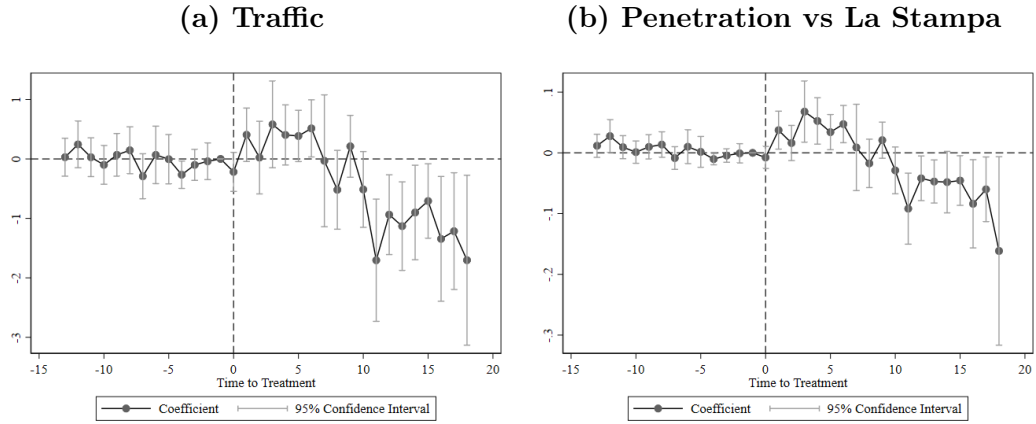
Notes: Balance check: Significance of distance in explaining the pre-June 2015 covariates. Estimates pertain to an OLS estimation of the covariate of interest on commuting distance to the border for the pre-policy subsample, accounting for time fixed effects (except for the census sourced, time-invariant covariates). For comparability, all variables are standardized. Standard errors are two-way clustered: by municipality and month-year.

Figure A.8. Local news searches vis à vis national news.



Notes: Proxy for local-news penetration. Values over time. This proxy is calculated as the ratio between internet searches for the national outlet i ($i \in \text{Il Fatto Quotidiano, La Repubblica, La Stampa, Il Giornale}$) and local sources s (appearing as different colors in the graphs), as retrieved via the Google Trends platform. Local sources whose Google Trends search did not report any result were assigned a value of zero and do not appear in the graphs.

Figure A.9. Effect on traffic and penetration

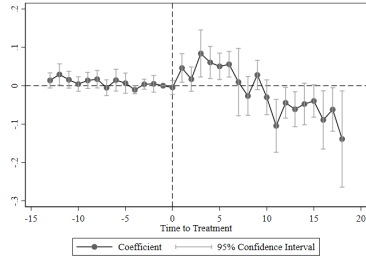


Notes: Coefficients plot from an event study specification. Local news traffic (left) and penetration with respect to national outlet La Stampa (right) are the outcomes of interest. Figure A.10 shows additional results for other National newspapers; namely: La Repubblica, Il Fatto Quotidiano, Il Giornale.

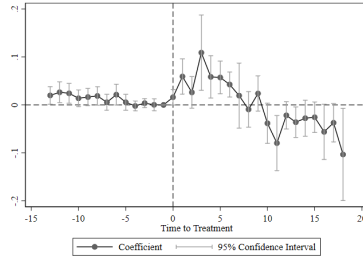
The figure suggests that, after the border controls, traffic and penetration of local news experienced an initial increase the closer to the border areas (though this increase is marginally insignificant for the traffic outcome). Over time, however, traffic is relatively higher further from the border, following the same direction of anti-immigrant slant. The magnitude of the negative coefficients in the later periods is larger than that of the positive ones in the initial periods and, on aggregate, the effect of proximity on traffic after the policy is negative, as documented by the last column of Table 2. The initial positive effect may be motivated by the presence of higher coverage, which is observed at the border and might have affected traffic positively. This effect, however, dynamically reverses and becomes negative, possibly due to anti-immigrant slant taking a delayed but prevailing role in explaining demand for local news.

Figure A.10. Effect on penetration, cont

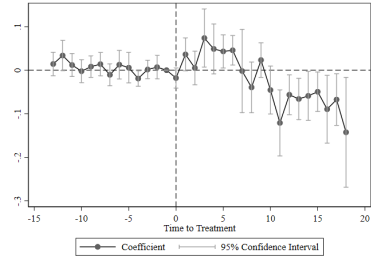
(a) vs La Repubblica



(b) vs Il Fatto Quotidiano



(c) vs Il Giornale



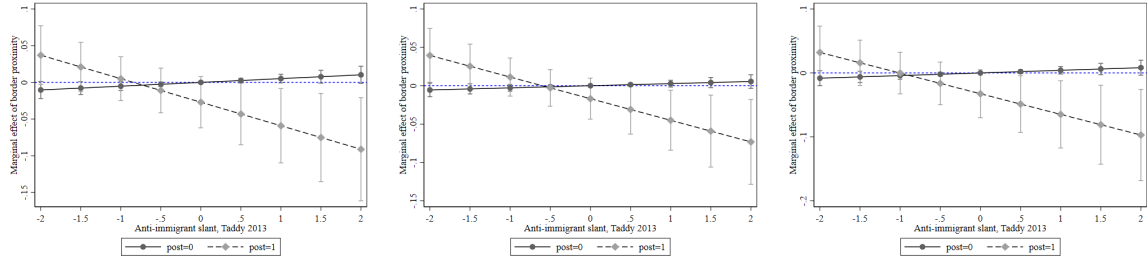
Notes: Coefficients plot from an event study specification. Local news penetration is the outcome of interest. From the top-left to the bottom-right, the national outlets of reference for the measure are, respectively: La Repubblica, Il Fatto Quotidiano, Il Giornale.

Figure A.11. Effects on traffic and penetration by levels of slant, cont

(a) vs La Repubblica

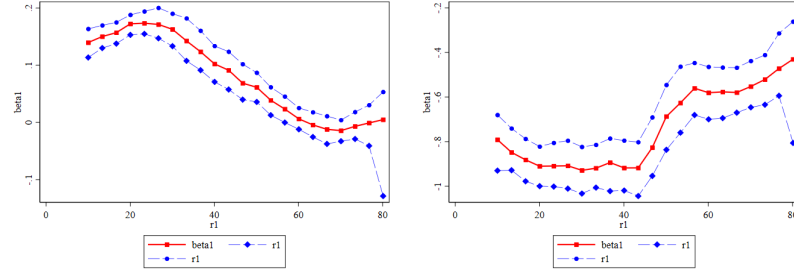
(b) vs Il Fatto Quotidiano

(c) vs Il Giornale



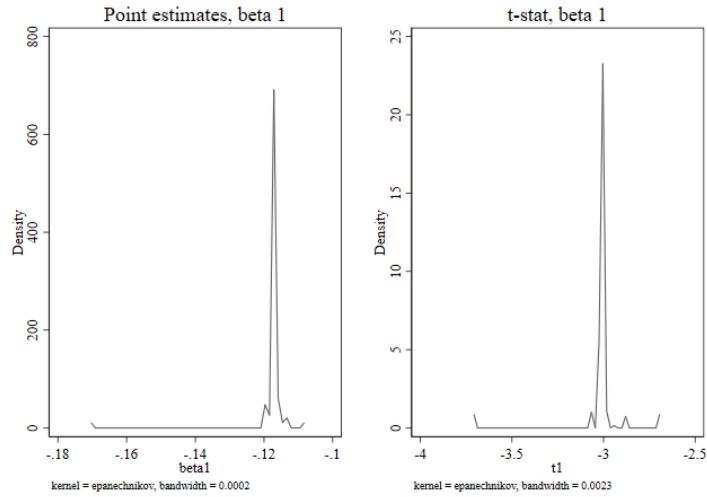
Notes: continuation of Table 8. Marginal effects from a triple-differences specification. The local-news penetration proxy is the outcome of interest. The national outlets of reference for the measure are, respectively: La Repubblica, Il Fatto Quotidiano, Il Giornale.

Figure A.12. β_1 robustness with alternative distance thresholds



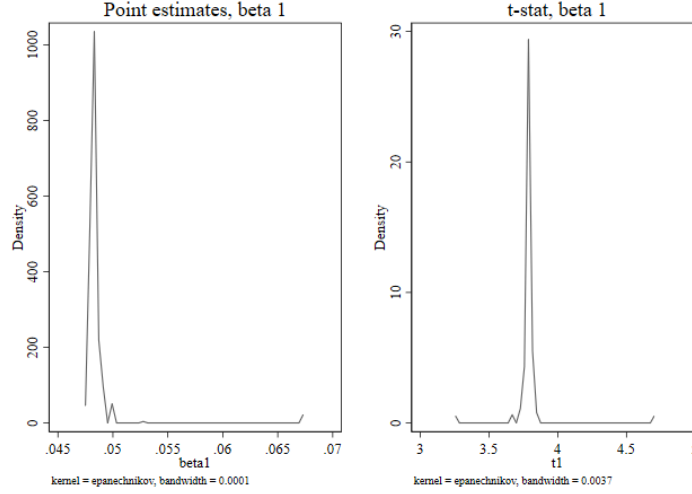
Notes: This figure presents the distribution of estimation results from a alternative specification that considers a binary treatment for distance: on the x-axis the threshold level of commuting distance up to which units are considered to be treated: e.g. for distance= n , treatment is defined as 1 if municipality m is at distance $>n$. The control sets are comparable to the baseline table's preferred column. Threshold values are plotted on the x-axis and the relative point estimate (red), with 95% confidence interval (blue, dotted lines) constitute the y-coordinates. On the left, graphs pertain to news-coverage. On the right, they refer to anti-immigrant slant estimations. P-values and confidence intervals are obtained with wild cluster-bootstraps based on 2000 repetitions. Clustering is one-way, on panel units.

Figure A.13. Coefficients robustness on news coverage: Leave-one-municipality-out estimates



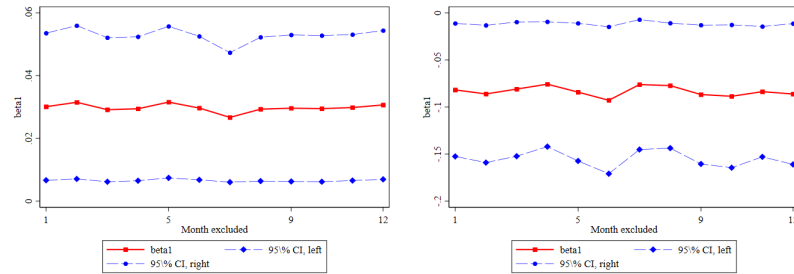
Notes: This figure presents the distribution of estimation results for β_1 , where, at each iteration, one municipality is left out of the sample. The reference specification is the baseline's preferred column. In the figure, the left-hand side depicts the distribution of point estimates for β_1 . The t-statistics distribution for these estimates is on the right.

Figure A.14. Coefficients robustness on anti-immigrant slant: Leave-one-municipality-out estimates



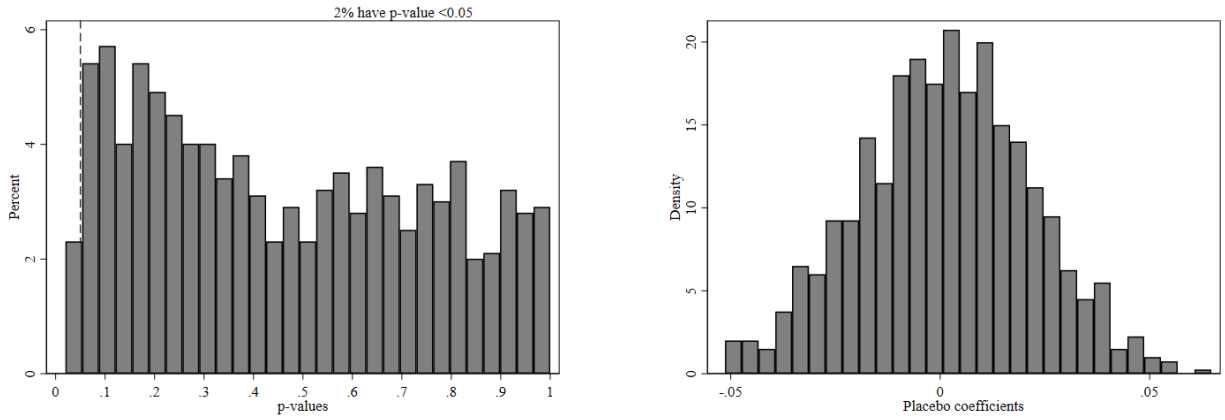
Notes: This figure presents the distribution of estimation results for the treatment coefficient, where, at each iteration, one municipality is left out of the sample. The reference specification is the baseline's preferred column. In the figure, point estimates for coverage are on the left and those for slant on the right.

Figure A.15. Coefficients robustness: Leave-one-month-out estimates



Notes: This figure presents the distribution of estimation results for the coefficient of interest. At each iteration, one month is left out of the estimation sample. The reference estimates are baseline specifications including all control variables. The point estimate (red), with 95% confidence interval (blue, dotted lines) constitute the y-coordinates. Graphs pertain to estimations on news coverage and anti-immigrant slant respectively on the left and on the right.

Figure A.16. Placebo predictions for slant



(a) Placebo, p-values

(b) Placebo, coefficients

Notes: Placebo estimates for slant. Slanted news items are classified by shuffling randomly the labels in the training sample. The distribution of p-values resulting from 1000 shuffling iterations is on the left, while coefficients appear on the right.

Appendix B. Additional Tables

Table B.1. Anti-immigrant slant: Selected examples of the results

Title	Source	Period	Label
<i>En.</i> , They sneak onto "commuter trains" to return to Italy after being expelled: two arrests <i>It.</i> , Si intrufolano sui "treni dei pendolari" per rientrare in Italia dopo l'espulsione: due arresti	riviera24.it	Jan 2012	1
<i>En.</i> , Giovanni Toti intervenes on the migrants matter: "We should not allow that Italy becomes a big refugee camp for Europe" <i>It.</i> , Giovanni Toti interviene sulla questione migranti: "Non si può consentire che l'Italia diventi un grande campo profughi dell'Europa"	sanremonews.it	Jun 2015	1
<i>En.</i> , Ventimiglia: police reports two migrants after a theft in Carrefour supermarket <i>It.</i> , Ventimiglia: la Polizia denuncia due migranti dopo un furto al supermercato Carrefour	Imperia Post	Nov 2018	1
<i>En.</i> , Call from the regional council to promote the diffusion of different cultures <i>It.</i> , Bando della Regione Liguria per promuovere la diffusione delle diverse culture	bordighera.net	Jan 2012	0
<i>En.</i> , Women, migration, and entrepreneurship, a congress at the Priamar <i>It.</i> , Donne, migrazioni e imprenditoria un convegno al Priamar	ivg.it	Mar 2015	0
<i>En.</i> , Ventimiglia: another 144 packed meals were delivered today by the Red Cross to migrants <i>It.</i> , Ventimiglia: altri 144 pacchetti alimentari consegnati oggi dalla Croce Rossa ad altrettanti migranti	sanremonews.it	May 2016	0

Notes: Examples from the predictions of Taddy (2013)'s model on the local news sample.

Table B.2. Summary Statistics

	Mean	Standard Deviation	N	Min	Max
Anti-immigrant slant, Friedman 2010	0.453	0.076	12636	0.182	0.654
Anti-immigrant slant, Taddy 2013	0.000	1.000	12636	-2.515	3.092
Anti-immigrant slant (predictions)	-0.001	1.001	12636	-2.818	2.865
Anti-immigrant slant (predictions), Taddy 2013	-0.000	1.000	12636	-2.389	2.906
Migrants' news exposure (log)	3.675	0.429	12636	2.079	5.101
Log of total news	7.600	0.238	12636	7.132	8.581
Avg. word count per outlet	464.513	66.535	12636	298.267	712.771
Population, log	7.155	1.325	12636	4.691	11.032
Population over 65, log	5.904	1.291	12636	3.466	9.789
Population density (population per sq. km)	235.963	342.012	12636	4.611	1922.043
Share of population over 65	0.291	0.051	12636	0.170	0.477
Stocks of immigrant population, log	4.673	1.386	12636	0.000	8.765
Inflows: New registrations	1.718	1.294	12636	0.000	7.650
Outflows: Cancellations	1.681	1.247	12636	0.000	5.832
(Taxable) Income per capita, log	9.957	0.131	12636	9.664	10.723
Commuting distance to Ventimiglia, minutes	45.755	20.075	12636	0.000	83.460
Total of reported crime events, yearly (Prov)	9.454	0.161	12636	9.210	9.740
Unemployment rate, 2011	0.079	0.027	12636	0.013	0.159
Share of university educated, 2011	0.074	0.025	12636	0.025	0.167
Max altitude	954.985	428.289	12636	157.000	2166.000
Area, sq. km	19.985	18.042	12636	1.289	100.658
Total voters, 2008	2370.596	5355.497	12636	67.000	39405.000
Top party is RW, 2008	0.840	0.366	12636	0.000	1.000
Shares LW coalition, 2008	0.322	0.069	12636	0.181	0.544
Top party is RW, 2018	0.425	0.494	12636	0.000	1.000
Shares RW coalition, 2018	0.464	0.068	12636	0.286	0.741

Notes: Summary statistics for the sample of interest.

Table B.3. Extended results: Heterogeneity

	News coverage (log)				
	(1) Benchmark results	(2) Education, 2011	(3) LW vote shares, 2008	(4) Altitude (max)	(5) Unemployment rates, 2011
Inverse distance (min)					
# Post	0.030** (0.012)	0.025* (0.013)	0.018 (0.015)	0.014** (0.006)	0.007 (0.021)
Triple Interaction		0.132 (0.199)	0.088 (0.092)	0.000*** (0.000)	0.293 (0.253)
Adjusted within R ²	0.22	0.22	0.22	0.25	0.22
	Anti-immigrant slant				
	(1)	(2)	(3)	(4)	(5)
Inverse distance (min)					
# Post	-0.083** (0.036)	-0.052 (0.036)	-0.041 (0.036)	-0.031** (0.014)	-0.028 (0.052)
Triple Interaction		-0.888** (0.423)	-0.324* (0.192)	-0.000*** (0.000)	-0.712 (0.558)
Observations	12129	12129	12129	12129	12040
Adjusted within R ²	0.08	0.09	0.09	0.11	0.08
Municipality FE	Yes	Yes	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes	Yes	Yes
Geo and News	Yes	Yes	Yes	Yes	Yes
Months to elections	Yes	Yes	Yes	Yes	Yes
Population controls	Yes	Yes	Yes	Yes	Yes
Migration controls	Yes	Yes	Yes	Yes	Yes
Income and crime	Yes	Yes	Yes	Yes	Yes
Prov. level trends	Yes	Yes	Yes	Yes	Yes

Notes: Coefficients relate to the dependent variable *news coverage* at the top, and *anti-immigrant slant* at the bottom. The set of controls included is equivalent to the rightmost of the benchmark table's columns. Column 1 is included for comparison and indicates benchmark results. Column 2 allows for an interaction with education levels (as of 2011). Column 3 considers initial political ideology (as from 2008's elections), Column 4 and 5 include, respectively, the average altitude of the municipality and the levels of unemployment as of 2011. Standard errors are two-way clustered. Stars correspond to the following p-values: * p < .10, ** p < .05, *** p < .001.

Table B.4. Time trends at municipality level

	Dep. variable:	
	(1)	(2)
	News coverage (log)	Anti-immigrant slant
Inverse distance (min)		
# Post	0.037**	-0.079**
	(0.014)	(0.039)
Observations	12129	12129
Adjusted within R ²	0.12	0.05
Municipality FE	Yes	Yes
Municipality time-trends	Yes	Yes
Month-year FE	Yes	Yes
Geo and News	Yes	Yes
Months to elections	Yes	Yes
Population controls	Yes	Yes
Migration controls	Yes	Yes
Income and crime		

Notes: Extended results, linear time trends interacted with municipality dummies. The table structure is comparable to the baseline estimates. Stars correspond to the following p-values: * p < .10, ** p < .05, *** p < .001.

Table B.5. Extended results: Distance functional forms

Linear distance		
	(1)	(2)
	News coverage (log)	Anti-immigrant slant
Linear distance # Post	-0.111*** (0.018)	0.320*** (0.062)
Adjusted within R ²	0.27	0.13
Log distance		
	(1)	(2)
	News coverage (log)	Anti-immigrant slant
Log distance, # Post	-0.073*** (0.016)	0.158** (0.055)
Adjusted within R ²	0.107	0.074
Quadratic distance		
	(1)	(2)
	News coverage (log)	Anti-immigrant slant
Distance, # Post	-0.069*** (0.017)	0.239*** (0.052)
Linear distance, squared # Post	0.027** (0.009)	-0.052** (0.018)
Observations	12218	12218
Adjusted within R ²	0.29	0.13
Municipality FE	Yes	Yes
Month-year FE	Yes	Yes
Geo and News	Yes	Yes
Months to elections	Yes	Yes
Population controls	Yes	Yes
Migration controls	Yes	Yes
Income and crime	Yes	Yes

Notes: Extended results, Different functional forms for the distance-based treatment. Results compare to the rightmost columns of the baseline results. Stars correspond to the following p-values: * $p < .10$, ** $p < .05$, *** $p < .001$.

Table B.6. Robustness to the exclusion of the imputed sample

	Dep. variable:	
	(1)	(2)
	News coverage (log)	Anti-immigrant slant
Inverse distance (min)		
# Post	0.009**	-0.051***
	(0.003)	(0.012)
Observations	3214	3214
Adjusted within R ²	0.21	0.04
Municipality FE	Yes	Yes
Month-year FE	Yes	Yes
Geo and News	Yes	Yes
Months to elections	Yes	Yes
Population controls	Yes	Yes
Migration controls	Yes	Yes
Income and crime	Yes	Yes
Tot. outlets		

Notes: Robustness test of the main results against the use of the non-imputed sample only. The dependent variables are the three media outcomes of interest. Standard errors are two-way clustered at municipality and month-year level. Controls are gradually added from left to right. Stars correspond to the following p-values: * $p < .10$, ** $p < .05$, *** $p < .001$.

Appendix C. The Political Context in Italy

As of 2021, Italy was formed by 7,904 municipalities (source: ISTAT), which constitute the lowest level of government. Municipalities are organized in 107 territorial bodies which correspond to 83 provinces, 14 metropolitan areas and 10 otherwise denominated entities, with an equivalent broad role. Provinces are then grouped into 20 regions. At the national level, the Italian Parliament is bicameral, composed of the Chamber of Deputies and Senate of the Republic. Elections of their members are held every five years.

The recent Italian political landscape is rooted in a major political crisis which took place during the early 90s, sparked by a series of scandals that were famously dubbed *Tangentopoli* (Bribeville). Specifically, a set of investigations uncovered extensive corruption systems that implicated entrepreneurs and politicians in power, for both major political parties at the time (the Italian Socialist Party and the Christian Democrats). In this setting, entrepreneur Silvio Berlusconi, although involved in some connected investigations, gained influence and formed a center-right coalition which allowed him to become the Prime Minister for the first time in 1994. This set of scandals marked the end of the so-called *First Republic* and the beginning of the *Second Republic* in 1994, which period saw the rise of a bipolar landscape with two major coalitions: the center-left and center-right. Several authors mark the 2013 elections as defining the end of this Second Republic (D’Alimonte, 2013; Durante et al., 2019; Tebaldi, 2014). In 2013 in particular, comedian Beppe Grillo brought his newly created Five Stars Movement to the elections, obtaining an unanticipated, relative success. Five Star gained significant traction among an unsatisfied, unrepresented electorate, and internet diffusion was found to have a major role in the movement’s ascent (Campante et al., 2018). This political movement was considered by some scholars as taking a broad populist stance by several sources (Bratti et al., 2020; Campo et al., 2021). Conditional on this scenario, and with the migration crisis peaking in 2014-2015, Italy saw the rise of two right-wing parties that maintained a strong nativist, anti-immigrant rhetoric (Bratti et al., 2020; Campo et al., 2021; Gamalerio and Negri, 2023; Romarri, 2020): Lega and Fratelli d’Italia (FDI). Lega’s party was founded in 1991 as a separatist party that broadly aimed at a secession of the more productive Northern Italy from the Center-South and therefore supported political and fiscal federalism. The party’s vote shares oscillated from around 8% in 1992’s general elections, to 4% in 2013, after joining Berlusconi’s coalition, to above 17% in 2018, under new party head Matteo Salvini. The new leader abandoned the Northist stance and maintained a strong anti-immigrant rhetoric. Lega’s Manifesto of 2018 condemned NGO operators in the Mediterranean Sea and stated that immigration should be based on cultural proximity. Fratelli d’Italia is a nationalist party that emerged in 2012 from a separation from the center-right coalition.

Giorgia Meloni took the lead since 2014. The party's identity is associated with a conservative, nationalist ideology of post-fascist heritage. On migration issues, it took an anti-reception position. FDI had been part of the same center-right coalition as Lega in the 2013 general elections. While failing to reach a 2% vote share in the 2013, 2018's general elections saw the party's electorate increase more than two fold, exceeding 4% of the vote. In the elections of 2022, the party's success peaked: it was the most voted in the right-wing coalition, leading Giorgia Meloni to become prime minister.

Appendix D. News sample collection

Appendix D.1. Validation

To better understand how representative is the news sample in the analysis, this section contains additional elements of the news data collection. It also assesses their relevance for the areas of interest.

Table D.1 lists all outlets that could be identified in the area of interest. The website homepage appears in the first column. Column 2 shows the first availability in the dataset. Column 3 summarizes whether that source is available fully, partially, or not at all. In the last column, a note on the availability is added if necessary. Figure D.1 shows the amount of Google Trends traffic for the sources that were not available in full. This traffic is plotted against one fully available source (Sanremonews), which is used as a benchmark (in purple). The graph shows that unavailable sources make up a very small percentage of the benchmark

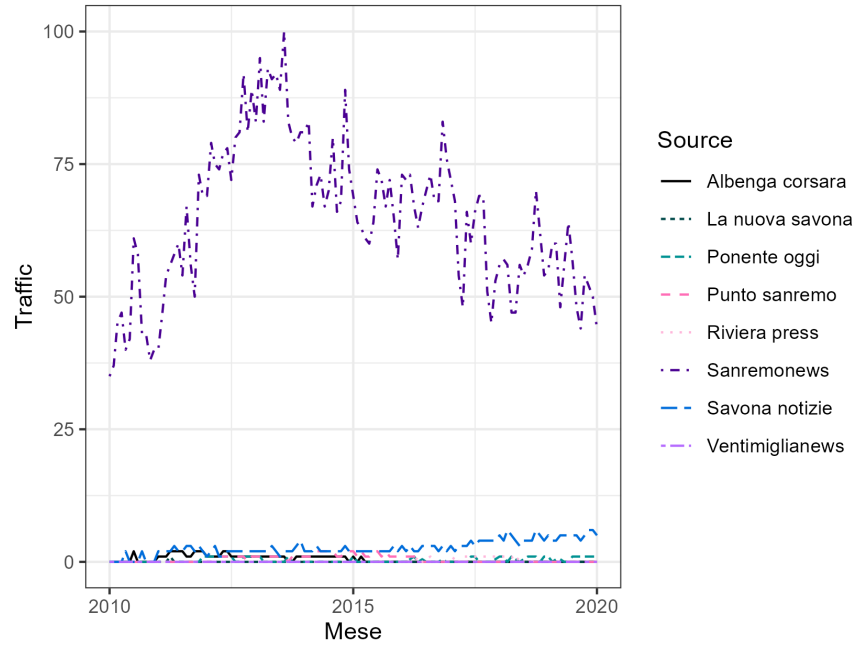
traffic (6.5% yearly on average).

Table D.1. List of outlets

Name	News availability	In data collection	Other notes
sanremonews.it	Since 2006	Yes: full period	
riviera24.it	Since 2008	Yes: full period	
rivierapress.it	Since 2016	Yes	Registered legally in Dec 2014. Discontinued source.
savonanews.it	Since 2004	Yes: full period	
rivieratime.news	Since 2017	Yes: full period	Registered legally in Dec 2016
imperiapost.it	Since 2013	Yes: full period	Registered legally in Mar 2014
imperianews.it	Since 2018	Yes: full period	Currently changed name to lavocediimperiam.it
ecodisavona.it	Since 2012	Yes: full period	
104news.it	Since 2016	Yes: full period	
ivg.it	Since 2008	Yes: full period	
bordighera.net	Since 2010	Yes: full period	Discontinued source.
albengacorsara.it	Since 2014	Yes	Registered legally in Oct 2009. Archive available since 2014.
bordigheratv.it	Since 2011	Yes: full period	Registered legally in Jan 2011.
primalariviera.it	Since 2016	Yes: full period	Online version of paper outlet <i>La Riviera</i> . No info on official date of registration. Full period claim is inferred based on Google Trends and Wayback-Machine.
ventimiglianews.it	Since 2018	Yes	Informative blog, non-regular frequency. Appears to exist since at least 2010 (twitter). Archive available since 2018
ponenteoggi.it	-	No	Registered legally in Apr 2011; 0 traffic in the sampled period from Google Trends
puntosanremo.it		No	Registered legally in 2012. Discontinued source. Last activity on Twitter: Feb 2018. Last homepage available in Wayback Machine: Feb 2019. Omitted because not available at collection.
lanuovasavona.it		No	Registered legally in Nov 2017. First article in archive on "migranti" in May 2018. Omitted because not available at collection.
savonanotizie.it		No	Existed between 2005 and the beginning of 2019. Then the same owners created redazionews.it. Omitted because not available at collection.

Notes: News outlets created after the study's period are excluded. The study also doesn't cover news outlets delivering news at wider geographies (regional/national) nor outlets from other media tools (tv, paper). Blogs/forums are omitted (with the exception of ventimiglianews which represents a small portion of data). It doesn't include outlets that have dedicated topics (sports/arts) outside the scope of analysis.

Figure D.1. Traffic for missing sources, given benchmark

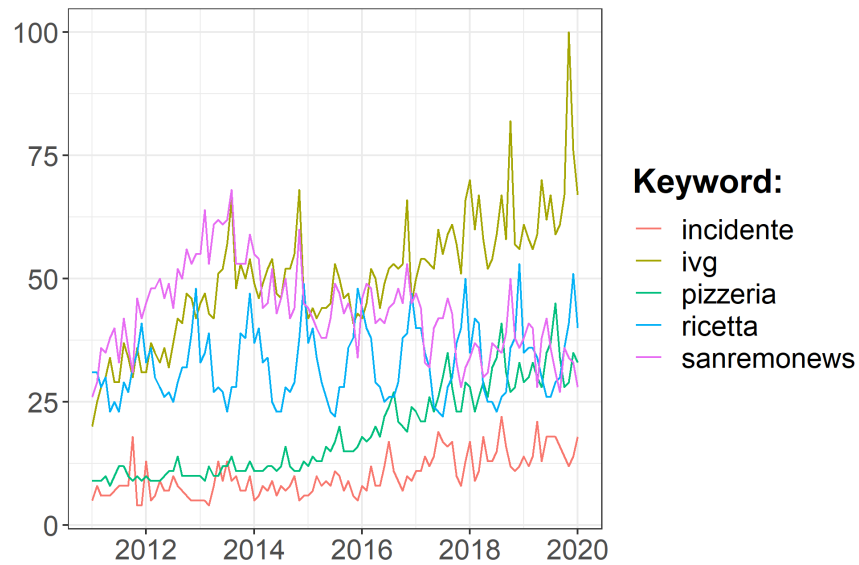


Notes: This plot shows the google traffic for news sources that end up missing, partially or totally, in the final dataset. Data are extracted from Google Trends and pertain to the region of Liguria, Italy. Sanremonews is appearing in the purple line and is not missing, but added for size comparison.

To understand the overall interest for the news in sample, Google Trends searches for the two most popular sources—*Sanremonews* and *IVG*—are contrasted with searches for a set of alternative, reasonably popular internet search terms—“accident” (It., “incidente”), “pizzeria,” “recipe” (It., “ricetta”)—for the region of Liguria. This is done to make sure that the readership size for the local outlets collected is of economic relevance. The search comparisons are plotted over time in Figure D.2. The graph shows that local news is very much comparable to this set of search items, which suggests that the local readership is substantial. Figure D.3 complements this information by plotting trends on internet usage in Liguria, compared with the rest of Italy, and based on ISTAT statistics.

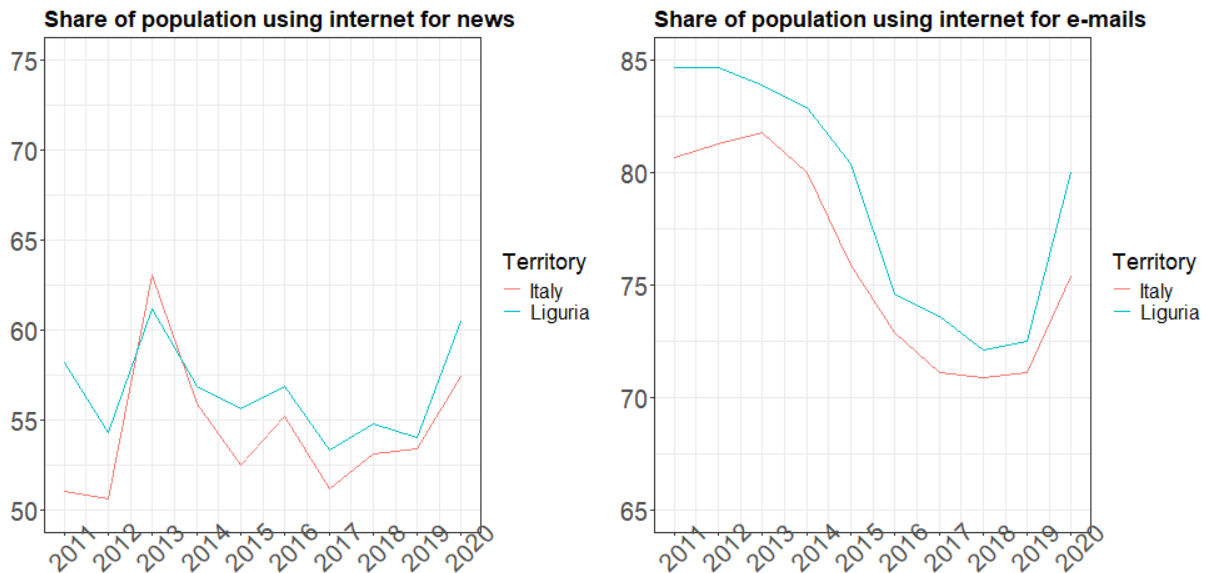
Figure D.2. Local news readership

Local news vs other searches



Notes: Comparison of Google searches for two of the main local news outlets, versus other internet searches.

Figure D.3. Internet usage for news consumption

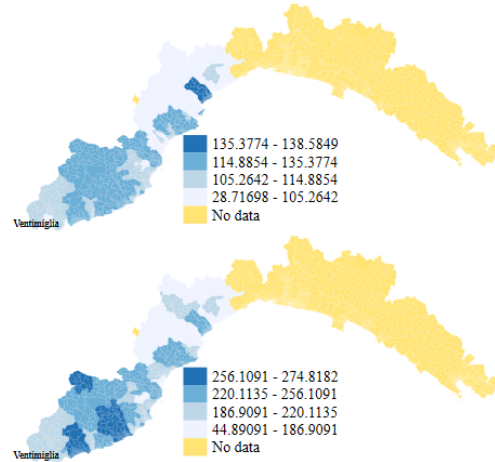


Notes: Share of population that uses the internet, sourced by ISTAT. On the left, is the share of the overall population that uses the internet for news. On the right, is the proportion of the population that uses the internet for e-mail.

Appendix D.2. Geocoding and Imputing News

A simple imputation method expands on the initial proxy to a set of municipalities not appearing in Google Trends. Municipality m_1 's information is imputed from its nearest neighbor municipality m_2 , where the neighborhood is established based on commuting patterns. Data on commuting patterns at the municipality level for 2011 is sourced by ISTAT, with information on how many individuals residing in municipality m_1 commute to work or study in municipality m_2 . If the percentage of commuters among the residents is less than 25%, data are not imputed.³⁶ Otherwise, information for municipality m_1 is proxied with the average across commuting destinations $m_k \in K$, where the set K is such that the difference between the share of all commuters and the share of commuters for which data on m_k is available is no more than 20%. Information for m_1 is imputed as the weighted average of information from observed js, where importance weights are the commuting shares. This procedure is reiterated and leads to data as in Figure D.4. Table F.1 lists all 132 municipalities for which news data has been geo-localized in both the periods before and after June 2015.

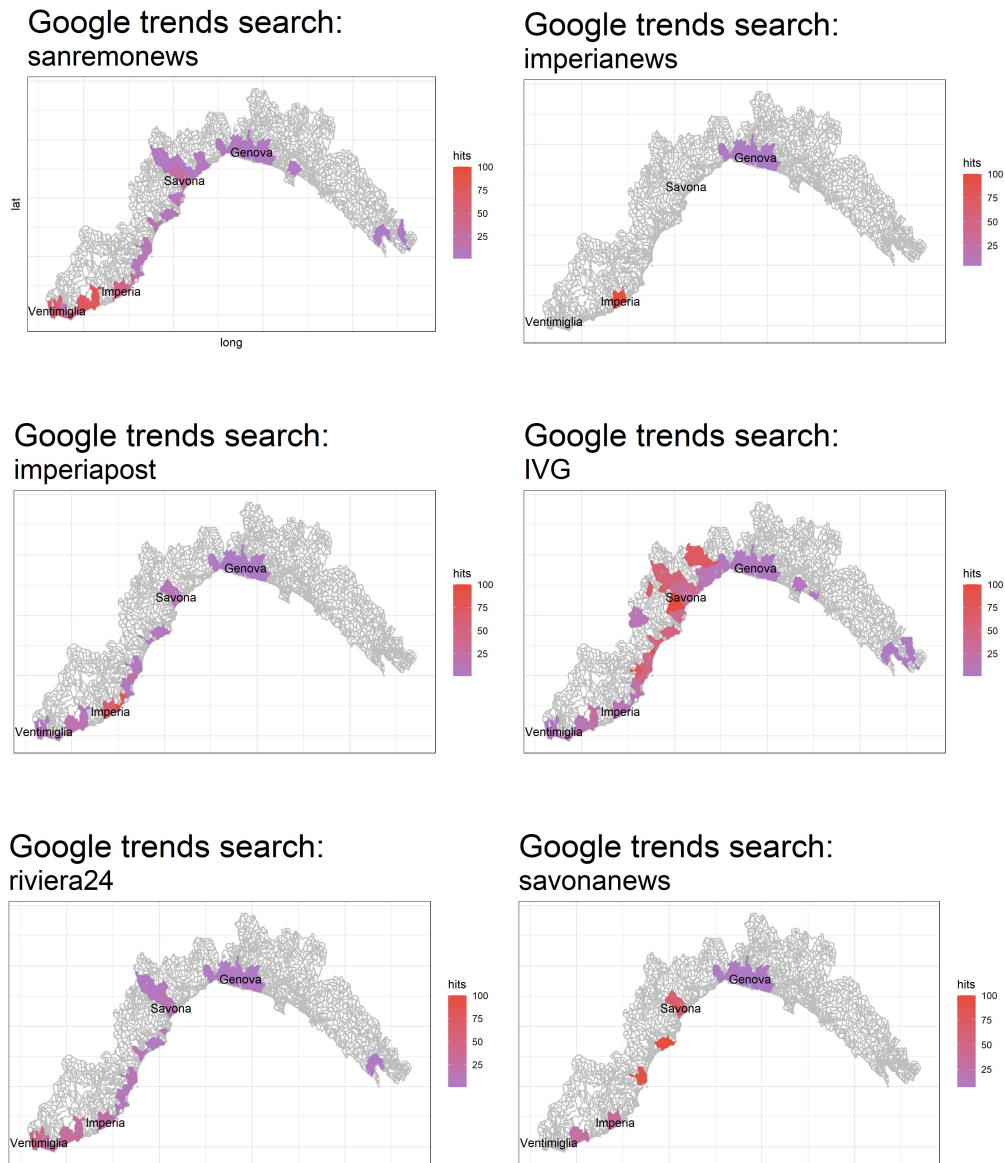
Figure D.4. News sample geographic distribution, before and after June 2015



Notes: Data visualization of the news sample. On the top map are data related to the pre-treatment period, while the post-treatment period is illustrated on the bottom. News coverage patterns result from the geo-localization process described in the paragraph and pertain to the provinces of Imperia and Savona, which are respectively the closest and second closest provinces to the French border in the region of Liguria.

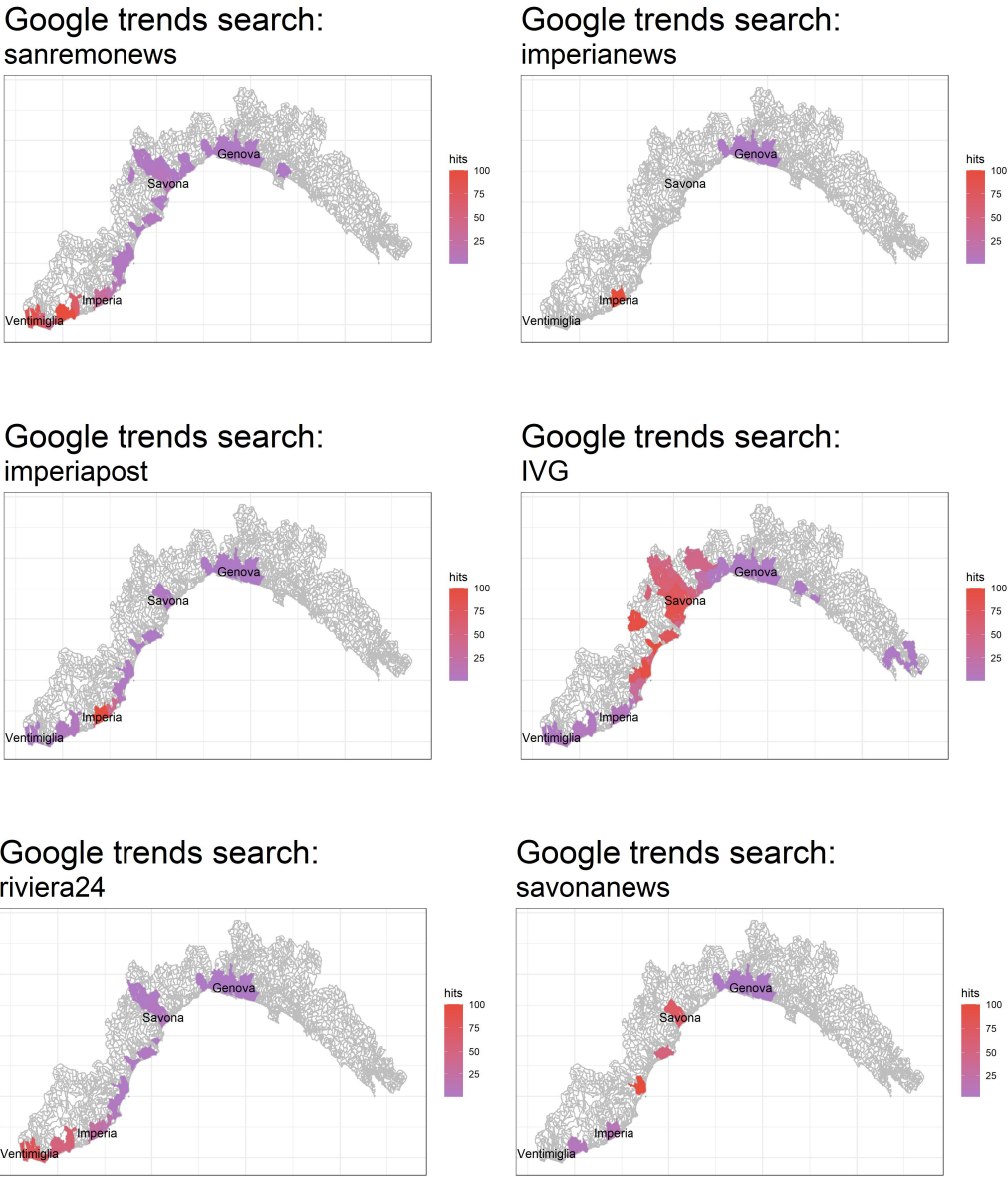
³⁶The rationale is that the higher the level of commuters, the more likely the news at the commuting destination may reach the commuters.

Figure D.5. Local news sources, Google trends data. Pre-treatment



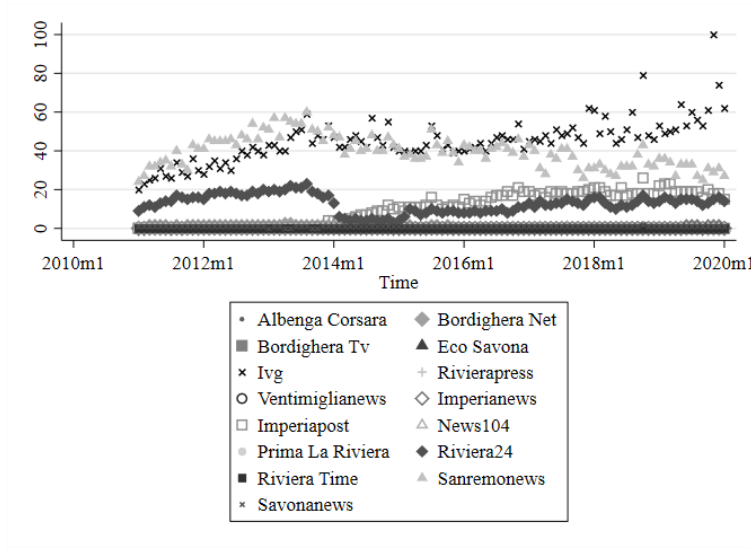
Notes: Local news sources and geographic interest of users, based on Google Trends data for the period ranging from 2010 to 2015.

Figure D.6. Local news sources, Google Trends data. Post-treatment



Notes: Local news sources and geographic interest of users, based on Google trends data for the period ranging June 2015 to 2020.

Figure D.7. Local news sources, Google Trends data. Traffic over time



Notes: Local news sources and time-varying interest of users, based on Google trends data for the period ranging from Jan 2011 to 2020.

Appendix E. Text-Based Measures

Appendix E.1. News cleaning

To obtain the final sample of migration-related news, extracted articles that only matched with the keyword “foreigner” were further screened. This step was implemented to avoid the presence of articles that would be, for instance, tourism-related and captured unintentionally with the use of this keyword. To clean this part of the corpus, a topic model was constructed via Latent Dirichlet Allocation (DTA), following the seminal contribution of Blei et al. (2003) and the R-package *topcmodels* by Hornik and Grün (2011). Topic modeling is a methodology that allows the classification of documents into categories along an unsupervised approach. LDA is arguably the most common of such methodologies, and involves a Bayesian mixture model that treats each article as a mix of a set of κ topics, and each topic as a mix of ω words. Differently from a clustering method that would strictly assign each article to one single group, here each news item is allowed to be assigned to several topics, with some proportion. The number of topics must be fixed a priori, it was set to 7 in this application³⁷. Let a corpus be formed of $i \in \{1, \dots, N\}$ documents, containing text that belongs to a dictionary of $j \in \{1, \dots, M\}$ terms. Each document i is formed by a subset of M_j terms.

Following Hornik and Grün (2011), the model specification assumes the following generative process:

³⁷A higher κ appeared to produce redundant topics

The term distribution β is determined for every topic by:

$$\beta \sim \text{Dirichlet}(\delta) \quad (\text{E.1})$$

The proportions θ of the topic distribution for the document i are:

$$\theta \sim \text{Dirichlet}(\alpha) \quad (\text{E.2})$$

For every j term,

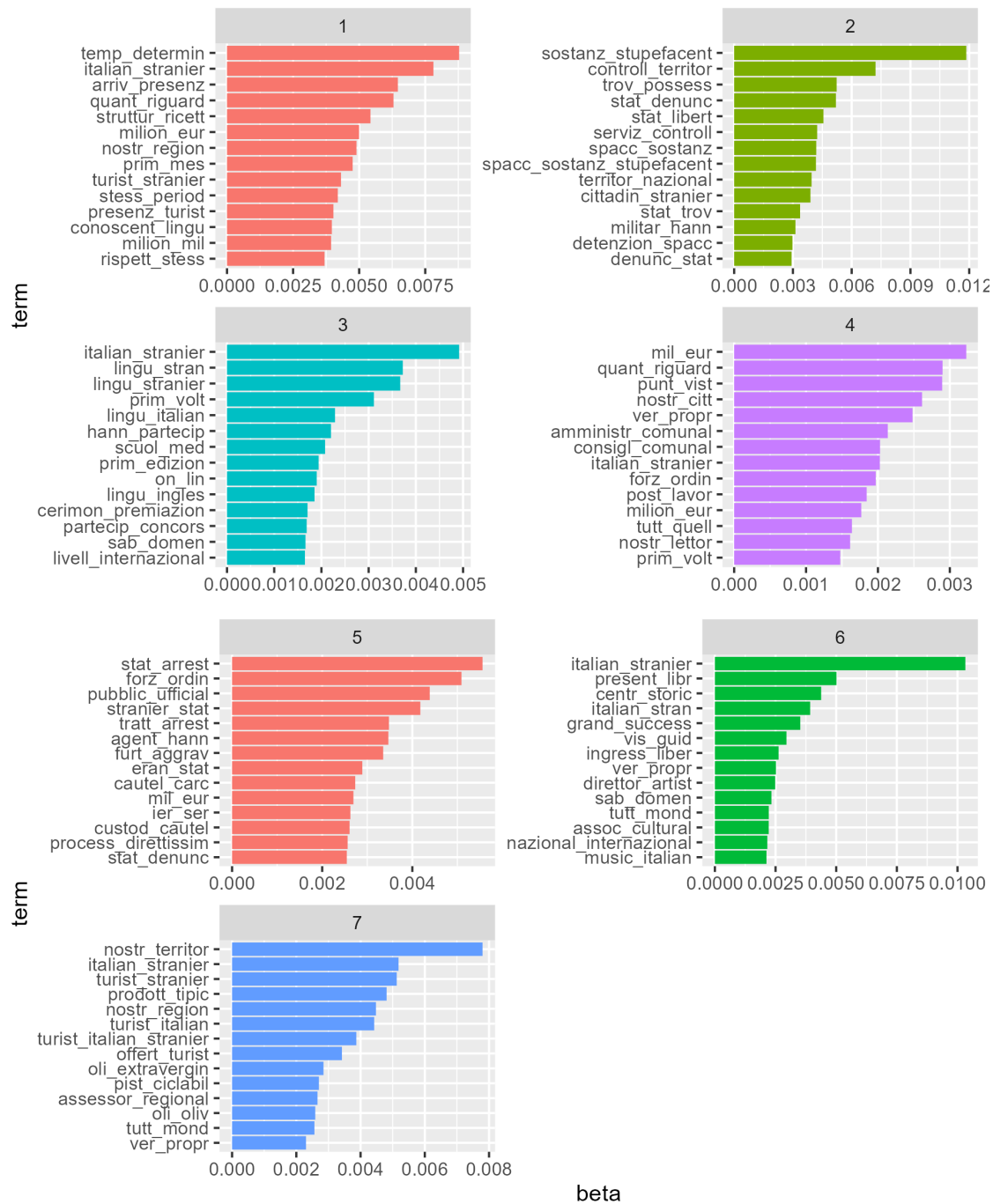
- Choose a topic $z_j \sim \text{Multinomial}(\theta)$
- Choose a word j from a multinomial probability distribution conditioned on topic $z_j : p(j|z_j, \beta)$.

The estimation methodology follows from a variational expectation-maximization (VEM) algorithm, where the alpha parameter is estimated (see Hornik and Grün (2011) for details). To summarize the resulting topic distribution, a plot of the most recurring terms for each topic is presented below in Figure E.1. From this visualization, the 7 groups can be identified as relating to the following subjects:

topic 1:	labor market (with tourism)
topic 2:	crime, drugs
topic 3:	events, competitions, ceremonies
topic 4:	local governance
topic 5:	crime prosecution
topic 6:	cultural events
topic 7:	tourism sector

As a decision rule, if the probability for document i to relate to topics 1, 3, 6, or 7 is above 90%, that document is excluded as these topics are the least likely related to migration (keep in mind that no other migration-related keyword other than "foreigner" is found for these articles). This process excludes 9,580 articles from the initial sample of news.

Figure E.1. LDA model, frequent terms by group.



Notes: Tables of most frequent terms in each topic distribution.

Appendix E.2. Media Slant

Among the prominent contributions in text-based measures of media slant, Groseclose and Milyo (2005) constitute a seminal example. The authors use US Congressional citations to estimate the political positions of think tanks and then use their mentions to predict media partisanship in twenty news outlets. Gentzkow and Shapiro (2010) build on this method and obtain media slant indices for 433 outlets, by enlarging the set of terms from sole think tank citations, to a broader set of text occurrences, selected in terms of relevance via a χ^2 statistic. Given this feature selection, the authors predict slant via a two-step supervised generative model. The broad idea behind Groseclose and Milyo (2005) and Gentzkow and Shapiro (2010) is to use text issued from politicians' discourse, whose affiliation is known, as a *training set* to predict the ideology of newspapers, unknown a priori but inferable from how similar it is to the political speeches. More recently, Taddy (2013) improves on these methodologies by proposing a model in which a multinomial inverse regression reduces the dimensionality of the predictors (terms) to a univariate score. A *forward regression* step then predicts the response (slant labels in this context), given this reduced information. In this application, Taddy (2013) is benchmarked against an alternative elastic-net penalized logistic regression (Friedman et al., 2009, 2010) and found to perform very similarly in a validation step. To construct the training set, this study relies on a set of four news sources whose slant (or lack thereof) is taken as evident. This constitutes a difference from Groseclose and Milyo (2005) and Gentzkow and Shapiro (2010), who employ politicians' speeches as their predetermined, labeled sample. Employing news to train the model guarantees that they belong to the same journalistic dimension as the unlabeled, target news set. Other than this comparability argument, a motivation for this choice is data availability. The use of Italian political speeches would present data availability limitations, further complicated by the complexity of a fractionalized political scenario that does not directly reduce to the more tractable bipolar *Democrats-Republicans* setting adopted in the US studies. A discussion of the Italian political landscape is detailed in Appendix Appendix C.

The classification proposed here involves the construction of a simple binary variable, where a value of 1 corresponds to the detection of similarity with anti-immigrant discourse and 0 to the lack of it. The *training sample* is constructed as follows.

Anti-immigrant labeled news items in the training sample derive from national newspapers *La Verità* and *Il Giornale*. The former outlet was founded in 2016 and openly takes an anti-immigrant stand. This is documented, for instance, in the online section discussing the migration crisis context, entitled *Chronicles of Invasion* (*it. Cronache dell'Invasione*). According to survey statistics diffused by the market research platform

YouGov,³⁸ this news outlet is considered right-wing by 81% of respondents. Of these, 17% classify it as far-right, the highest percentage among all news outlets in the list. 12% of people classify it as center-oriented or non-oriented, while the remainder consider it a left-wing source. Data from *La Verità* are collected from its online platform by considering all news that is included in the *Chronicles of Invasion (it. Cronache dell’Invasione)* section, and complement this extraction with migrant-related news by a similar keywords-search fashion as for my main dataset. The latter source *Il Giornale* shares similar Yougov reviews with *La Verità* and also takes critical rhetoric against migration.³⁹ Data from the *Il Giornale* are collected from the online platform by considering all news from the same keyword searches used for the local news sample. These news articles take label value 1.

Non-right-wing labeled news in the training sample derives from two sources. One is the website *L’Unità news*, self-defined as a not-for-profit news aggregator of *antifascist and anti-racist information*. The website also hosts the archives for and shares the name with the discontinued newspaper *L’Unità* founded by Antonio Gramsci in 1924 as the official news outlet for the Italian Communist Party. News published in this online platform are extracted, following the keywords-search fashion as for the main dataset. The second source is the online version of the regional newspaper *Il Secolo XIX*. This outlet was founded in Genova, Liguria in 1886. Since 2015, the outlet had been affiliated with the national source *La Stampa* and since 2017 been part of the editorial group GEDI, which includes, among others, the daily *La Repubblica* and the weekly *L’Espresso*.⁴⁰ News collection for *Il Secolo XIX* follows the same keyword-based steps. Articles coming from these two sources are labeled with a 0 in the binary slant measure.

Via this procedure, the training set is constituted of 5926 0-labeled and 5293 1-labeled instances, which will serve to classify the main news dataset. This study proposes two classification approaches: i) a penalized maximum likelihood logistic model with elastic-net regularization (Zou and Hastie, 2005; Friedman et al., 2010), with 10-fold cross-validated tuning parameters λ (the penalty parameter) and α (the level of mixing between lasso and ridge, with $\alpha = 1 \implies$ full lasso). This is compared with ii) the alternative Taddy (2013)’s

³⁸<https://it.yougov.com/news/2019/05/03/giornata-mondiale-della-liberta-di-stampa/>

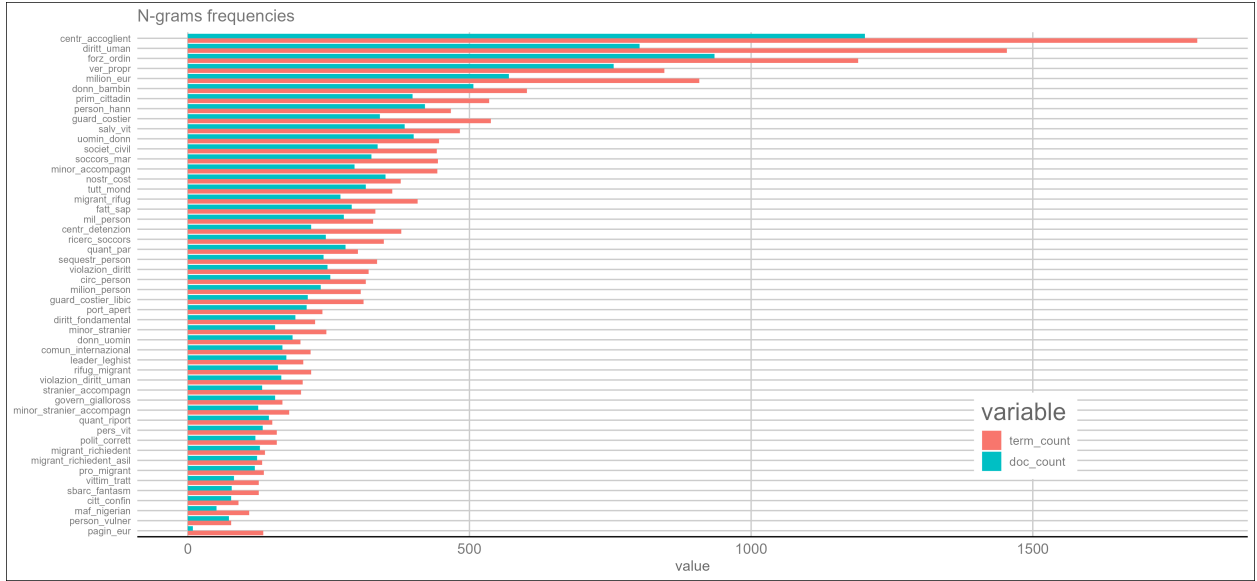
³⁹An headline example from February 2022: <https://www.ilgiornale.it/news/politica/carola-leroina-fuori-legge-2007541.html>: “Carola Rackete, the hero of the radical-chic, rightwing world, reveals that she was even more Taliban than the reception of the German NGO Sea Watch.”

⁴⁰According to the aforementioned YouGov report, *La Stampa* is perceived as center-oriented by 33% of respondents and left-wing by 39%. *La Repubblica* is considered left-wing leaning by 79% of the interviewed.

methodology.⁴¹

For every article in the training set, a document-term matrix is constructed as described in section 3.3, with rows constituting the training set articles i and columns constituting the n-grams j ($n \in 2, 3$) present in the articles. Pre-processing steps are performed via the R package *text2vec*⁴². The vocabulary of n-grams is pruned via a feature selection based on the χ^2 statistic of Gentzkow and Shapiro (2010). For an idea of the matrix content, Figure E.2 summarizes the frequencies of some top frequent n-grams.

Figure E.2. Document-term matrix, sample of frequent terms.



Notes: Frequency table of most frequent terms from the training sample.

In Friedman et al. (2010)'s method, the document-term matrix cells will contain the metric $tf - idf_{ij}$. the objective function for logistic regression⁴³ is the penalized negative binomial log-likelihood function, given N sample size and p parameters:

$$\min_{(\beta_0, \beta) \in \mathbb{R}^{p+1}} - \left[\frac{1}{N} \sum_{i=1}^N y_i (\beta_0 + x_i^T \beta) - \log(1 + e^{(\beta_0 + x_i^T \beta)}) \right] + \lambda [(1 - \alpha) \|\beta\|_2^2 / 2 + \alpha \|\beta\|_1] \quad (\text{E.3})$$

On the right, the penalty severity is determined by the tuning parameter λ . This regularization is a convenient approach for models with a high number of parameters, which may suffer from over-fitting problems. The elastic net mixes between two penalty functions: the ridge-regression penalty ($\alpha = 0$) and the lasso-regression penalty ($\alpha = 1$). The motivation behind this mixing lies in the aim to exploit the qualities of

⁴¹The two measures share the same prediction in roughly 86.7% of instances.

⁴²<https://CRAN.R-project.org/package=text2vec>

⁴³Given an observed binary outcome $Y \in 0, 1$, the logistic model can be written as $\frac{p(Y=1|X=x)}{p(Y=0|X=x)} = \beta_0 + x^T \beta$.

each approach while overcoming their limitations if considered singularly. lasso allows for a better parsimony of the model as it shrinks some parameters to zero. In case of high correlations between predictors, however, lasso may omit relevant information, as would only select one coefficient, ignoring the rest. Friedman et al. (2010) suggest that $\alpha = 1 - \epsilon$ for small $\epsilon > 0$ may already be a convenient solution to avoid lasso degeneracy and conserve its sparsity property. In this study, hyper-parameters λ and α are chosen from a 10-fold cross-validation procedure on the training sample. Figure E.1 displays the levels of α (on the left) and λ (on the right) that minimize the misclassification rate (on the y-axis).

Table E.1. Tuning parameters for the Elastic net.

alpha	lambda
0.100000	0.009858

Notes: Cross-validation results. Best tuning parameters.

Binary classification returns probability scores, where a probability corresponds to the estimated degree of likelihood that the article is labeled as 1 (= right-wing). In the benchmark, for a probability score higher than 0.5, the article is assigned the value 1 (some variation on this threshold is considered in the extended results).

In Taddy (2013)'s procedure, for each document i , equipped with term counts $\mathbf{x}_i = [x_{1i}, \dots, x_{Ji}]'$, the n -gram total counts for labels $y \in \{0, 1\}$ are defined as $x_y = \sum_{i:y_i=y} \mathbf{x}_i$. Given row sums $m_i = \sum_{j=1}^J m_{ij}$, the multinomial inverse regression would then be:

$$\mathbf{x}_y = MN(\mathbf{q}_y, m_y) \text{ with } q_{yj} = \frac{\exp[\alpha_j + y\phi_j]}{\sum_{l=1}^J \exp[\alpha_l + y\phi_l]}, j = \{1, \dots, J\}, y \in \{0, 1\} \quad (\text{E.4})$$

Where MN is a j -dimensional multinomial distribution with size $m_y = \sum_{i:y_i=y} m_i$ and probabilities $q_y = [x_{y1}, \dots, x_{yJ}]'$. Under a set of conditions, a sufficient reduction score for frequency $\mathbf{f}_i = \mathbf{x}_i/m_i$ is defined as $\mathbf{z}_i = \phi' \mathbf{f}_i, \implies y_i \perp\!\!\!\perp \mathbf{x}_i$. Given this score, the forward regression would simply be a univariate model, which in the present case, takes the logistic form: $E[Y] = \frac{1}{e^{\beta_0 + \beta_1 z_i}}$. The author proposes a so-called *gamma-lasso* algorithm to find a joint maximum a posteriori estimate of both coefficients and their prior scale, by maximizing the log-likelihood of the model with a lasso penalty, with λ_j possibly varying with j . Further details on this estimation are in the author's contribution. Although the two alternative models produced very similar labeling, a set of validation steps leads to the adoption of Taddy (2013)'s measure as the baseline slant index. Figure E.3 represent some predictive terms, with associated scores.

Figure E.3. Most predictive coefficients, Taddy 2013.



Notes: Examples of terms used in Taddy (2013) classification. Positive x values indicate higher term impacts in absolute terms on the slant predictions.

Given this binary classification of documents, average indicators are computed at the month-year level for each municipality. These aggregate averages then can be interpreted as a proportion of right-slanted news out of migrant-related news to which a particular municipality was exposed. Importantly, given the news geo-localization described in the previous section, a document (news piece) can be read in more municipalities. When constructing these averages, importance weights were used according to the Google Trends-based information, as detailed in section Appendix D.2.

Appendix E.3. Media Slant Validation and Exploration

This section explores the validation procedures undertaken for the construction of the anti-immigrant slant index.

To maximize the prediction accuracy for the labelling process, models were cross-validated on the pre-labeled sample. A 10-fold cross-validation procedure was performed on both the baseline and the alternative anti-immigrant slant index. Table E.2 below presents cross-validation results. Generally, the two measures appear to be fairly comparable: the elastic-net measure appears in the first row and has an estimated accuracy of 0.85. The accuracy for Taddy's measure, in the second row (i.e., the proportion of correct classifications over the total sample size) is estimated to be 0.87. Both models present higher levels of specificity than sensitivity. Both values are above 80% for both measures. All in all, these statistics appear to favor the use of

Taddy (2013)’s index over the Elastic-net, as a benchmark indicator.

Table E.2. Cross-validation results

model	cv_acc	cv_acc_sd	cv_sens	cv_fscore	cv_spec
Elnet	0.88	0.01	0.88	0.88	0.88
Taddy	0.92	0.01	0.91	0.92	0.92

Notes: Performance results from the two classification models, based upon a 10-fold cross-validation, done manually for Taddy (2013). Columns present, from left to right, the following statistics: cross-validated accuracy (*cv_acc*) and its standard deviation (*cv_acc_sd*); sensitivity (*cv_sens*), F-score (*cv_fscore*) and specificity (*cv_spec*).

As a further measure of validation, results were also compared with human classification of roughly 900 random entries. This sample was extracted from the main corpus, with source-based stratification. The same text was analyzed by two *Amazon Mechanical Turk* annotators. Having multiple annotators for one task is key to capturing response variation, which may also reflect subjective biases in the response. Interestingly, this procedure revealed some evidence that human classification of slant is likely heavily influenced by predetermined ideology in a general crowd: human classifiers did not agree on slant in 27 % of the instances. Below, the full instructions entered in the platform are provided.

ORIGINAL TEXT:

Instructions: Per favore, legga il testo qui sotto elencato. Le si chiederà di interpretare se il contenuto della news ha un’accezione politica. Recentemente, il dibattito politico in Italia si è incentrato molto sul tema della migrazione. In generale, i partiti a destra sostengono un’ideologia critica riguardo ai recenti sbarchi dei richiedenti asilo, e del loro arrivo in Italia, affermando per esempio problemi sicurezza emergono per i cittadini. Il centro sinistra si è visto più propenso all’accoglienza. Le viene domandato se l’articolo afferma, anche generalmente/indirettamente/secondariamente, il tema della migrazione. Per esempio se l’articolo riporta un furto ad opera di individui stranieri presenti in Italia, il tema della migrazione, per quanto indirettamente, è presente. Se l’articolo parla soltanto di eventi cultura o di turismo, in quel caso non affronta il tema della migrazione. Le sarà poi chiesto di valutare se da chi ha scritto l’articolo si evince un tono vicino ad un’ideologia politica (e quale). Per esempio: -se l’autore sceglie il termine “nullafacenti” per descrivere dei cittadini stranieri/migranti, si evince un tono critico sulla migrazione, e quindi maggiore vicinanza al discorso di destra. -Se l’articolo giudica importante o necessario un evento di accoglienza organizzato nel Comune di riferimento, questo tono è più vicino al discorso di sinistra. Infine le viene domandato se il SOGGETTO dell’articolo, di cui il testo parla, ha un’inclinazione ideologica e quale. Per esempio, se il testo menziona la decisione di un sindaco di non approvare un progetto di accoglienza nel nome di una maggiore sicurezza dei cittadini, il soggetto dell’articolo si categorizza di un’ideologia più a destra. (NOTA:La vicinanza politica dell’autore può differire da quella del soggetto del testo). Un’ultima sezione le chiede un breve commento sulla sua scelta, basta la citazione di un punto del testo, una parola chiave o una sua breve frase che confermi che le risposte sono inserite a seguito di una sua valutazione.

Testo dell'articolo:[...]

Questo articolo parla, anche generalmente, di migrazione? Risponda con SI o NO.

I fatti descritti nell'articolo sono delineati neutralmente o con un'accezione ideologica? Risponda con N se neutralmente, e con O se pensa esista del contenuto ideologico.

Il tono di questo articolo ha una somiglianza con il discorso politico di destra/centro-destra, di sinistra/centro-sinistra o nessuno dei due? Risponda con una delle tre opzioni: DESTRA; SINISTRA o NEUTRO.

Questo articolo menziona l'opinione o i fatti di una persona (un politico o cittadino o un intervistato) di destra/centro-destra, di sinistra/centro-sinistra o nessuno dei due? Risponda con una delle tre opzioni: DESTRA; SINISTRA o NEUTRO.

Indichi una frase o parola del testo che motivi la sua scelta. Basta una breve risposta. Se l'articolo non parla di migrazione (risposta 1), può inserire NOMIG.

ENGLISH VERSION:

Instructions: Please read the text below.

You will be asked to interpret if the news content has a political meaning. Recently, the political debate in Italy has focused a lot on the issue of migration. In general, the parties on the right support a critical ideology regarding the recent landings of asylum seekers, and their arrival in Italy, stating for example security problems arise for citizens. The center-left has seen more inclination to their reception. You are asked if the article affirms, even generally / indirectly / secondarily, the theme of migration. For example, if the article reports a theft by foreign individuals present in Italy, the theme of migration, albeit indirectly, is present. If the article speaks only of cultural or tourism events, in that case it does not address the issue of migration. You will then be asked to evaluate if the writer of the article reveals a tone close to a political ideology (and which one). For example: -if the author chooses the term "deadbeat" to describe foreign citizens/migrants, there is a critical tone on migration, and therefore greater proximity to the right-wing discourse. -If the article judges a welcome event organized in the relevant Municipality to be important or necessary, this tone is closer to the left-wing discourse. Finally, you'll be asked if the SUBJECT of the article, of which the text speaks, has an ideological inclination and which one. For example, if the text mentioned a mayor's decision not to approve a reception project in the name of greater citizen safety, the subject of the article is categorized with a more right-wing ideology. (NOTE: The political closeness of the author may differ from that of the subject of the text).

The last section asks you for a brief comment on your choice, you just need to quote a point in the text, a keyword or a short sentence from it to confirm that the answers are inserted after your evaluation.

Article text:[...]

Does this article also talk about migration, even generally? Answer with YES or NO.

Are the facts described in the article outlined neutrally or with an ideological meaning? Answer with N if neutral, and with O if you think there is ideological content.

Does the tone of this article bear a resemblance to right / center-right, left / center-left political discourse, or neither? Answer with one of the three options: RIGHT; LEFT or NEUTRAL.

Does this article mention the opinion or facts of a person (a politician or citizen or an interviewee) on the right / center-right, left / center-left or both? Answer with one of the three options: RIGHT; LEFT or NEUTRAL.

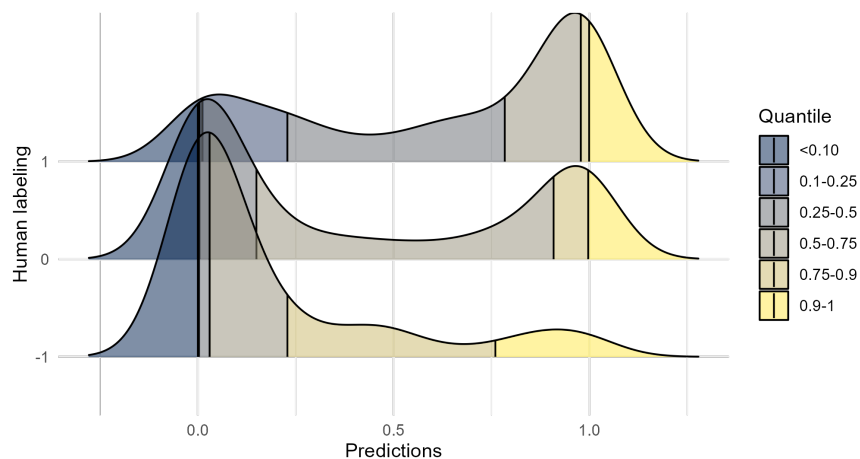
Indicate a sentence or word of the text that motivates your choice. A short answer is enough. If the article does not talk about migration (answer 1), you can enter NOMIG.

The questionnaire allowed for a different answer for two potentially different cases. 1) the author of the article wrote with some ideological inclination; 2) the subject of the article involves a person's opinion (or

facts) with a political tendency. For the scope of this study a broad definition of media bias is adopted: no distinction between the two cases is investigated, and giving voice to third persons with a political orientation is considered on the same level of writing with that orientation (recall, as broadly similar approaches, Groseclose and Milyo (2005) who define media-slant based on think-tank citations; Gentzkow and Shapiro (2006) estimating informative words via a data-driven approach, Taddy (2013)’s refinement; Flaxman et al. (2016) who propose a non-text-based measure, based on the geographic distribution of readership and voters rather than on text; Djourelova (2023) who focuses on the banned use of the expression “illegal immigrant”).

Slant indices and human classification agreed in 66% of cases. Figure E.4 plots the distribution of the baseline classification’s predicted probabilities, against the evaluation of human annotators. This last takes values 1 if at least an annotator coded the article as "right-wing", 0 if the article was coded neutral (or the two annotators gave opposite marks), -1 if the annotator assigned the article to be "left-wing" oriented. Both neutral and left-wing marked articles have their higher mass at low predicted probabilities.

Figure E.4



Notes: Densities of predicted probabilities in the benchmark classification (density functions), against human classification (y-axis factors).

All in all, these results point to a general validity of the adopted measures for slant. While the general behavior is as expected, measurement error exists, and several approaches to circumvent it are taken in the main analysis.

Appendix F. Municipalities in Sample

Table F.1. List of municipalities in the baseline panel

Municipalities:	Airole; Alassio; Albenga; Albisola; Albissola; Altare; Andora; Apricale; Aquila D'arroschia; Armo; Arnasco; Aurigo Badalucco; Bajardo; Balestrino; Bardineto; Bergeggi Boissano; Bordighera; Borghetto D'arroschia; Borghetto Santo Spirito; Borgio Verezzi; Borgomaro; Bormida Cairo Montenotte; Calice Ligure; Calizzano; Camporosso; Caravonica; Carcare; Carpasio Casanova Lerrone; Castel Vittorio; Castelbianco; Castellaro; Castelvecchio Di Rocca Barbena Celle Ligure; Cengio; Ceriale; Ceriana; Cervo; Cesio Chiusanico; Chiusavecchia; Cipressa; Cisano Sul Neva; Civezza; Cosio D'arroschia; Cosseria; Costarainera Dego; Diano Arentino; Diano Castello; Diano Marina; Diano San Pietro; Dolceacqua; Dolcedo; Erli; Finale Ligure Garlenda; Giustenice; Giusvalla; Imperia; Isolabona; Laigueglia; Loano; Lucinasco Magliolo; Mallare; Mendatica; Millesimo; Mioglia; Molini Di Triora; Montegrosso Pian Latte; Murialdo; Nasino; Noli Olivetta San Michele; Onzo; Orco Feglino; Ortovero; Osiglia; Ospedaletti Pallare; Perinaldo; Piana Crixia; Pietra Ligure; Pietrabruna; Pieve Di Teco; Pigna; Plodio Pompeiana; Pontedassio; Pontinvrea; Pornassio; Prelà; Ranzo; Rezzo; Rialto; Riva Ligure; Roccavignale; Rocchetta Nervina San Bartolomeo Al Mare; San Biagio Della Cima; San Lorenzo Al Mare; Sanremo; Santo Stefano Al Mare; Sassello; Savona Seborga; Soldano; Spotorno; Stella; Stellanello Taggia; Terzorio; Testico; Toirano; Tovo San Giacomo; Triora; Vado Ligure; Vallebona; Vallecrosia; Varazze; Vasia Vendone; Ventimiglia; Vessalico; Vezzi Portio; Villa Faraldi; Villanova D'albenga; Zuccarello
------------------------	---

Notes: Final list of Municipalities in the panel, resulting from the news geo-localization.

Appendix G. Supplementary Material

Supplementary material can be found in Peracchi (2025)