

BUSINESS CYCLE AND REALIZED LOSSES IN THE CONSUMER CREDIT INDUSTRY

Walter Distaso, Francesco Roccazzella,
Frédéric Vrins

LIDAM Discussion Paper LFIN
2023 / 07

LFIN

Voie du Roman Pays 34, L1.03.01 B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: lidam-library@uclouvain.be

<https://uclouvain.be/en/research-institutes/lidam/lfin/publications.html>

Business cycle and realized losses in the consumer credit industry*

Walter Distaso^{1,5,6}

Francesco Roccazzella^{2,†}

Frédéric Vrins³

¹ADIA, UAE

²IESEG School of Management, Univ. Lille, CNRS, UMR 9221

- LEM - Lille Economie Management, F-59000 Lille, France

³UCLouvain, LIDAM - LFIN, Belgium

⁵Imperial College Business School, United Kingdom

⁶UNIME, Italy

December 12, 2023

Abstract

We study the determinants of the loss given default (LGD) of consumer credit. Exploiting a dataset including more than 6 million of Italian consumer loans from 2007 to 2019, we find that macroeconomic and social variables significantly enhance forecasting performance both at the individual and portfolio levels, by up to 10 percentage points in terms of R^2 . This result is robust across forecasting exercises and model specifications. In particular, non-linear forecast combination schemes relying on neural networks are among the best performers in terms of mean absolute error, RMSE, R^2 , and model confidence set in every considered exercise. The relationship between the expected LGD and the macro predictors unveiled by accumulated local effects plots confirms the intuition that lower real activity, increasing cost-of-debt to GDP ratio, and greater economic uncertainty are associated with a greater LGD for consumer credit.

Keywords: Credit Risk; Consumer Credit; Loss Given Default; Non-Performing Loans

*The authors are grateful to the participants of the XVIII Credit Scoring and Credit Control Conference (Edinburgh, 08/30-09/01, 2023) for valuable feedback and discussions.

[†]Corresponding author - Email: f.roccazzella@ieseg.fr

1 Introduction

Consumer credit refers to the debt that individuals contract to purchase goods and services. This includes the usual credit lines (issued by credit institutions or by sellers directly) as well as liabilities resulting from unscheduled deferred payments.¹

Consumer credit is a fast-growing industry worldwide. As of September 2022, it reached USD 4,700 billion in the US and EUR 715 billion in the euro area. It accounts for 10.8% of the outstanding credit for households in the euro area.² Given the market size, studying the drivers of realized losses in the case of default is essential for creditors and all the stakeholders involved in the recovery process. Confronted with a non-performing credit (NPC), a debt holder typically has four main options: take her loss, initiate a collection procedure with the customer, give a mandate to a specialized firm that will collect the debt in her name, or sell the NPC at a discount to a company who will then try to recover part of the EAD by chasing the customer. To evaluate the relevance of initiating a collection procedure, determine the fees of a third-party debt collection firm, or to price the NPC to be sold or bought, every stakeholder needs to estimate what the recovery rate (RRs) are likely to be.

In the event of a default, the severity of the loss suffered by creditors can be measured using two parameters: the exposure at default (EAD) and the RR (or, equivalently, loss given default, $LGD = 1 - RR$). The EAD refers to the outstanding principal amount of the debt at the time of default, while the RR denotes the ratio of the amount that creditors will recover to the EAD, hence, takes value in $[0, 1]$. The EAD is typically known at the time of default. In contrast, the RR is only revealed when closing the debt collection process.

Although a vast literature deals with the RRs of defaulted corporate bonds, less is known about the RRs of consumer credits. One can hardly extrapolate the findings related to corporate bonds to consumer loans as the two types of debt are not easily comparable. To name just a few factors, the subjects involved and the empirical distributions of RRs are very different.³

When it comes to highlighting the main drivers of RRs, most of the existent literature has documented the central role of contract-specific variables (such as the seniority or the sector for bonds, and the maturity or the principal for consumer credits). Conversely, the impact of macroeconomic and social (MS) indicators on RR is less clear and seems to depend on the type of debt at hand. For instance, it has been known since the seminal paper of Altman et al. (2005) that realized recovery rates of corporate bonds are negatively correlated with the default rate. Surprisingly enough, this would not apply to consumer credit, where macroeconomic variables have largely been shown to be empirically irrelevant (see for example, Nazemi et al., 2022).

Different factors might have contributed to these findings. First, it is very difficult to access a meaningful consumer credit dataset: data is typically private, often scarce, and limited to a specific

¹The ECB defines consumer credit in the Bank lending survey for the euro area as "[...] loans granted for mainly personal consumption of goods and services. Typical examples of loans in this category are loans granted for the financing of motor vehicles, furniture, domestic appliances and other consumer durables, holiday travel, etc. Overdrafts and credit card loans also typically belong in this category. Consumer credit and other lending to households also include loans to sole proprietors and partnerships [...]. Loans included in this category may or may not be collateralized by various forms of security or guarantee" (See page 1 European Central Bank, 2016).

²Data on credit extended to individuals for household, family, and other personal expenditures, excluding loans secured by real estate for U.S. is available at <https://www.federalreserve.gov/releases/g19/about.htm> and for the euro area can be retrieved from the MFI balance sheets available from the ECB Statistical Data Warehouse <https://sdw.ecb.europa.eu/>.

³While the empirical distribution of the RRs of defaulted corporate bonds is smooth and mostly unimodal (Bellotti et al., 2021), the distribution of NPCs exhibits a pronounced bimodal pattern (see, e.g., Gambetti et al., 2019; Nazemi et al., 2022).

sector of the consumer credit industry, preventing the findings from being easily generalized. Second, the samples under analysis are relatively short in the time dimension while economic and social indicators are low-frequency time series. The low variability of MS indicators in short samples may explain their marginal relevance for describing consumer credit RRs compared to more granular credit-specific features such as exposure at default or debtor’s age. Third, research has focused on RRs at individual credit levels, despite defaulted consumer credits are often evaluated and managed at a portfolio scale. For instance, third-party collectors acquire the right to recover for *pools* of defaulted credits, earning the totality (if purchased) or a fee (if handled in the name of the debt owner) on the amount recovered. In both cases, they aim at maximizing the total recovered amount and, therefore, studying the determinants at the portfolio level better reflects the business model of this industry. Moreover, while MS indicators may have little predictive power at the individual level, they may play a bigger role at the pool level, where idiosyncratic risk is diversified.

Our research sheds some light on the aforementioned issues and makes several contributions to the literature on consumer credit. First, we jointly investigate the factors influencing consumer credit recovery rates originating from multiple industries, specifically from utilities, banking, financial, and commercial sectors. A database encompassing over 6 million NPCs in Italy, covering the period from 2007 to 2019 makes this comprehensive analysis possible. To the best of our knowledge, the dataset at hand is the most extensive examined in academic literature to date. Second, we perform our analysis on two levels: at the individual NPC level and at a portfolio level. While previous research primarily concentrates on the individual level, it is important to consider the portfolio level as well, for the reasons explained above. Third, we exploit the heterogeneity of economic dynamics across regions and provinces of Italy to understand how the economy impacts LGDs in the consumer credit industry. In contrast to previous studies focusing on specific types of NPC and the macroeconomic environment at the national level, we find that excluding MS indicators from the set of predictors significantly deteriorates the forecasting performance, both at individual credit and portfolio levels. Our findings are robust to the choice of forecasting models; they remain valid when considering standard regressions, machine learning algorithms, and combinations of those. Fourth, we find the use of artificial neural networks as a nonlinear combination strategy particularly encouraging since it attains the lowest MAE and RMSE values. Fifth, the accumulated local effect (ALE) plots confirm that our results are in line with economic intuition. For instance, a deterioration of the credit conditions at the national level is associated with higher LGDs for portfolios of NPCs.

The paper is structured as follows. In Section 2, we review the salient results related to the determinant of RRs for various types of debts. Section 3 provides an overview of our data. The methodology for sampling the data is exposed in Section 4 and Section 5 outlines the forecasting methods used in our study. Section 6 presents our results, and the final section concludes.

2 Literature review

In this section, we review the literature on the factors influencing RRs, with a particular focus on the significance of MS predictors. The empirical evidence on the importance of MS indicators for explaining consumer credit RRs is sparse and largely inconclusive. Results seem to differ depending on the timing of the studies, the category and sector of the NPC, and the type of MS variables taken into account.⁴

⁴A similar debate holds for non-performing loans: Dermine and de Carvalho (2006) find on a dataset covering the period 1995-2000 that GDP does not matter, in contrast with Bellotti et al. (2021) who find by analyzing a 1988-2015 dataset that

For example, Bellotti and Crook (2012) focused on defaulted credit card accounts (over the period 1999-2005) and found that including macroeconomic variables may enhance the predictive power of recovery rates. They observed that higher interest rates and unemployment levels at the time of default are associated with lower recovery rates, while higher earnings growth may lead to better recoveries. However, they also underline that their training sample size (374 defaults) was relatively small, and may not have been large enough for a more significant relationship between the economy and RRs to be discovered. Leow et al. (2014) found that macroeconomic variables may improve the predictive performance of mortgage loan RR estimates, but, conversely, not of personal loan RR. They considered a broad set of economic variables (including net lending growth, disposable income growth, GDP growth, net lending growth on dwellings, unemployment rate, saving ratio, interest rate, and the Halifax House Price Index) with data on defaulted mortgages (from 1990 to 2002) and unsecured personal (from 1989 to 1999). The legal environment also plays an important role. In the credit card industry, for instance, Fedaseyeu (2020) found that consumer protection legislation governing third-party debt collection reduces the number of third-party debt collectors and increases the LGD on delinquent credit card loans.

Beck et al. (2017) examined the RRs of consumer credits for goods and services in the period 2004-2008 in Germany, and focused on idiosyncratic determinants such as the EAD or prior debtor-specific collection rates, complemented by GDP growth and the regional unemployment rate. The latter variable was found to have a consistently negative effect on the RR. Nazemi et al. (2022) considered 65,535 defaulted unsecured consumer credits purchased between 2010 and 2013 from a German telecommunications company. In their case, including macro variables such as the provincial unemployment rate and excessive indebtedness provincial rate (the percentage of adults with strong negative credit ratings) did not lead to improved prediction accuracy.

3 Data

We consider three kinds of data. The first is a proprietary and anonymised database (referred to as NPC DB hereafter) owned by a third-party collector in Italy containing a broad range of consumer credits, including banking loans, financing contracts, sales credits, and debts related to telecommunication and utilities. The second are MS variables collected from various relevant sources, aiming to capture the effects of business cycles and the social environment. The third is a synthetic dataset about portfolios of NPC built from the first, aiming to represent the information that would be made available to a debt collector purchasing a pool of loans.

3.1 The Non-performing credits database

The NPC DB contains 6,493,794 NPCs from defaulted consumer credits that the third-party collector is entitled to recover for a maximum period of 365 days since receiving the mandate from the original lender. The data covers the period from January 2007 to December 2019. We have access to the debtor's postal code and the date when the third-party agency was authorized to recover the defaulted credit to match each NPC to a specific region and province at a specific moment in time. The database offers a representative picture of the situation in Italy since all 20 regions and 101 out of 107 provinces are represented.

macro indicators are key.

Table 1: Recovery rate descriptive statistics of non-performing credits by debtors type, sectors, and conditional on Total to Recover (TtR) ranging from EUR 50 to EUR 10,000.

Debtor	# obs	μ	σ	Percentile				
				.10	.25	.50	.75	.90
Female	3,001,281	0.25	0.4	0	0	0	0.5	1
Male	2,070,420	0.25	0.4	0	0	0	.5	1
Professionals	9,930	0.35	0.46	0	0	0	1	1
Unspecified	1,412,163	0.27	0.41	0	0	0	.622	1
Sector	# obs	μ	σ	.10	.25	.50	.75	.90
Telco	3,118,743	0.19	0.37	0	0	0	0	1
Commercial	504,604	0.31	0.43	0	0	0	.874	1
Utility	2,275,620	0.27	0.42	0	0	0	.723	1
Banking	240,936	0.48	0.47	0	.396	1	1	1
Financing	353,891	0.47	0.37	0	0	.479	.886	1
TtR	# obs	μ	σ	.10	.25	.50	.75	.90
50-100	764,869	0.42	0.48	0	0	0	1	1
100-250	1,933,739	0.31	0.44	0	0	0	.909	1
250-500	1,787,211	0.21	0.37	0	0	0	.323	1
500-1,000	1,128,216	0.18	0.35	0	0	0	.13	.96
1,000-2,500	621,764	0.16	0.33	0	0	0	0	.909
2,500-5,000	190,501	0.14	0.31	0	0	0	0	.777
5,000-10,000	67,494	0.15	0.32	0	0	0	.004	.855

Table 1 reports the descriptive statistics of the RRs across types of debtors and sectors, conditional on the Total To Recover (TtR) ranging from EUR 50 to EUR 10,000. Among the debtors, 78.10% are private individuals and a small fraction (about 0.15%) are professionals. The defaulted credits originate from utility (35.04%) and telecommunication sectors (48.03%), financing (car loans, credit cards, leasing, 5.45%), banking (overdrafts and personal credits, 3.72%) and commercial credit (sales credit, 7.78%). NPCs originating from the telecommunication industry (Telco) display the lowest average recovery rate of 19%, followed by the utility and commercial sectors, yielding 27% and 31%, respectively. NPCs originating from the banking and financing industry display higher RRs (respectively, 48% and 47%).

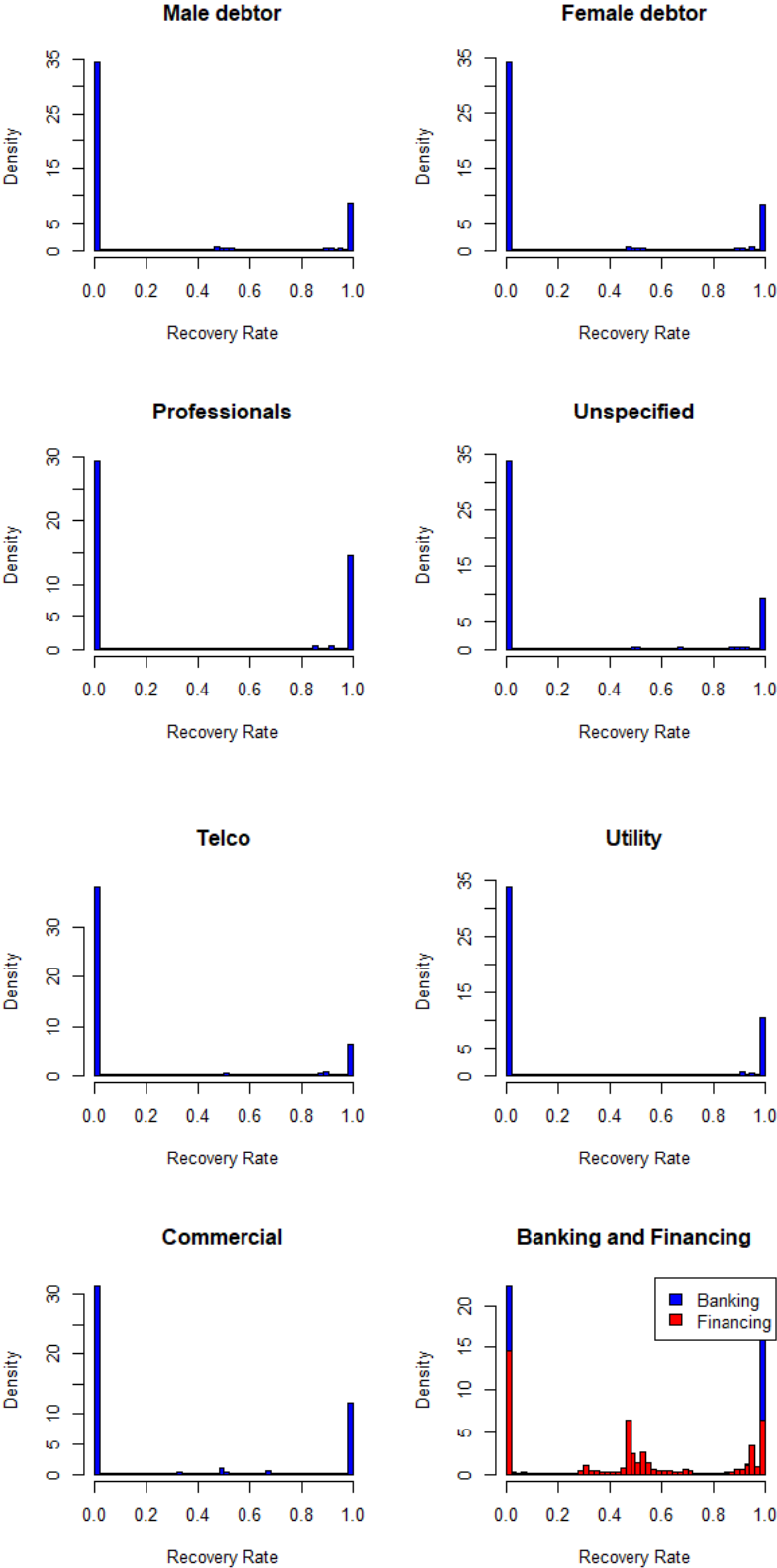
Figure 1 confirms the usual bi-modal distribution of recovery rates, with no-recovery ($RR = 0\%$) and full recovery ($RR = 100\%$) being the two most likely outcomes of the collection process. The Bernoulli-shaped distributions emphasize the high uncertainty inherent to consumer credit recovery rates. The only exception is the banking and financing sectors, which exhibit a third probability mass peak at 50%. The last section of Table 1 confirms previously documented evidence about the negative relationship between RRs and TtR (Nazemi et al., 2022).

We identify four categories of debtors: male and female individuals, professionals, and unspecified. We also observe the age of individual debtors. Each debtor is associated with a particular province and region, and each loan to a specific point in time. Hence, we can examine how NPC-specific predictors vary across Italy. To better explain the heterogeneity of the data across regions and highlight possible trends, we divide the regions into three macro-territories: northern, central, and southern.⁵

The top charts of Figure 2 display the average RRs (left), the TtR amounts (center), and the average ratio of principal-to-TtR amounts (right) for consumer NPCs in various Italian provinces, while those

⁵We follow the classification of the Italian National Institute of Statistics (ISTAT), which classifies the eight administrative regions of Aosta Valley, Piedmont, Liguria, Lombardy, Emilia-Romagna, Veneto, Friuli-Venezia Giulia, and Trentino-Alto Adige as the North, while Lazio, Marches, Tuscany, and Umbria are classified as the Center. The South includes the islands of Sicily and Sardinia, as well as Abruzzo, Apulia, Basilicata, Calabria, Campania, and Molise. More information available at: <https://www.istat.it/en/archivio/227202>.

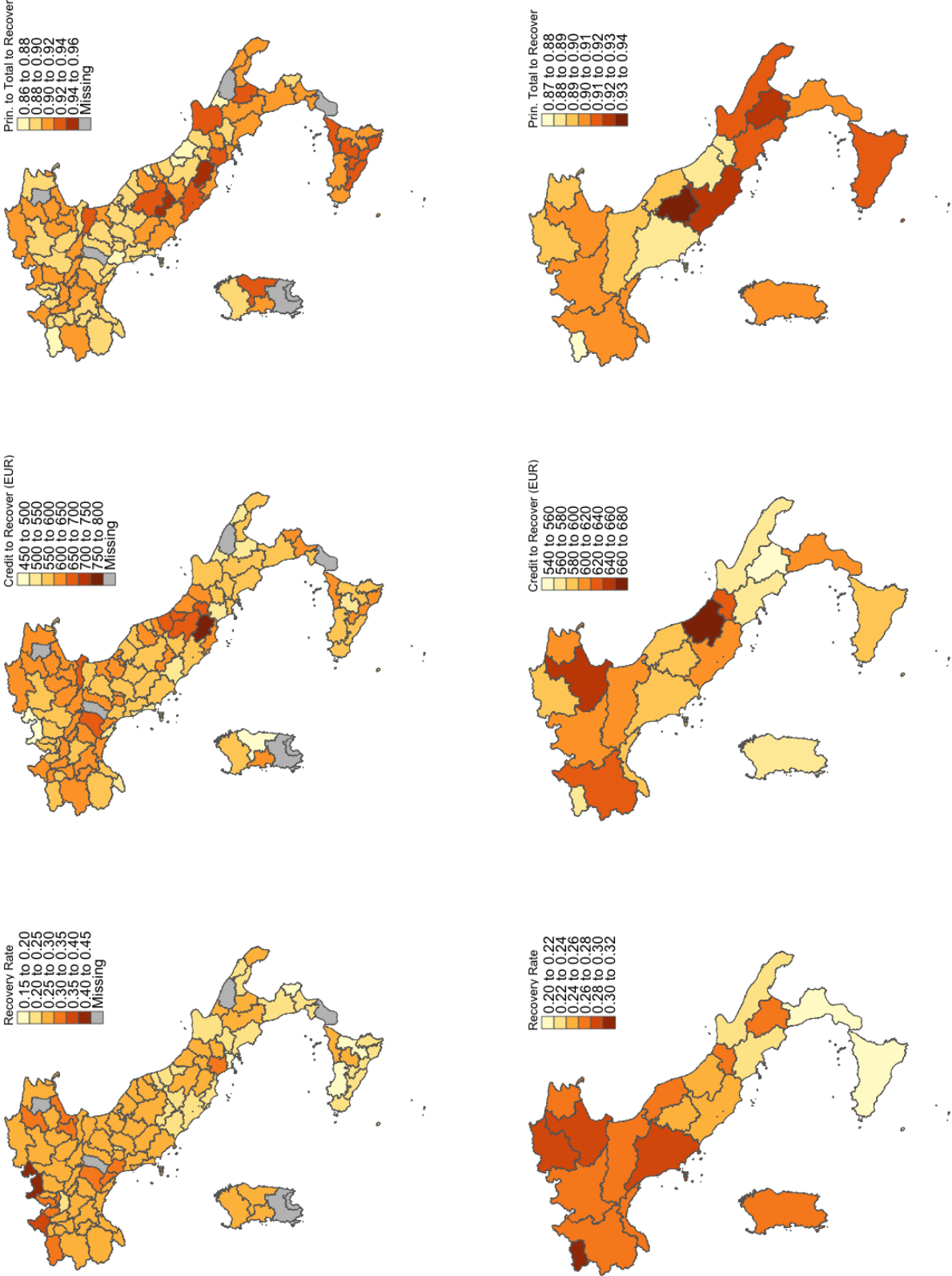
Figure 1: Distributions of the realized recovery rate of individual NPCs by debtors type and sectors. They display a strongly bi-modal shape, except in the banking and financing sector.



at the bottom show the same information but at the regional scale. Generally, the recovery rates for

NPCs are on average higher in the northern regions of Italy (average 28%, with standard deviation 42%) compared to the center (average 26%, with standard deviation 41.6%) and southern regions and islands (average 22%, with standard deviation 39.9%). However, there is still a significant amount of variability at the provincial level. In contrast to the differences in recovery rates observed between the north and south, the central panel of Figure 2 displays no consistent pattern for the total amount to recover, which is on average EUR 604 in the Center and North and EUR 550 in the South. The right panel indicates that interest and ancillary fees only contribute marginally to the total amount to be recovered, with the principal amount accounting for 86% to 96% of the total.

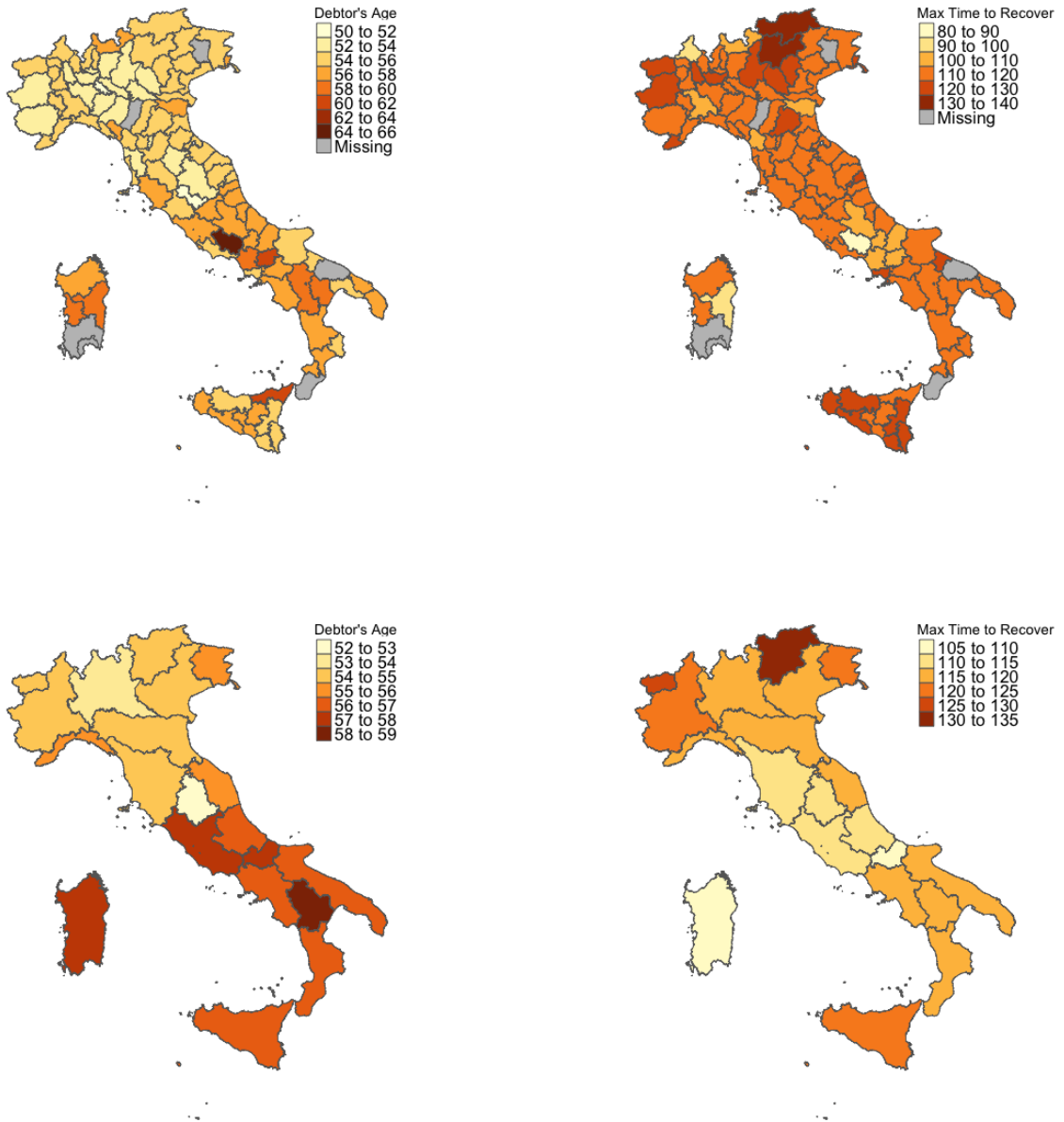
Figure 2: Heterogeneity of RRs (left), TtR amounts (center) and principal-to-TtR ratios (right) at province (top) and region (bottom) level.



This is particularly evident for NPCs originating from the center and south of Italy, where the principal accounts for over 90% of the total.

Figure 3 presents information on the average age of debtors (left panel) and the average maximum duration granted to third-party collectors to recover defaulted debt (right panel). We observe that debtors from the center and south of Italy have an average age of over 55 years, with some provinces such as Frosinone (Lazio), Caserta, Naples (Campania), and Messina (Sicily) having peaks at 60-66 years. We can also see the average maximum duration granted to third-party collectors to recover defaulted debt. The data shows that this duration is balanced across provinces and regions, with no significant trend emerging across macro-territories.

Figure 3: Heterogeneity of debtor's age (left) and maximum duration collection process (right) at province (top) and region (bottom) level.



3.2 Macroeconomic and social predictors

We use the debtor’s postal code and the date when the third-party agency was authorized to recover the defaulted credit to match each NPC to a specific region and province at a specific moment in time. This allows us to extend the contract-specific predictors discussed in the previous section with MS indicators. We have a total of 13 such variables that we can classify into four broad classes: business cycle, credit cycle, economic uncertainty, and social environment.

We start with the four business cycle indicators. We consider monthly seasonally adjusted unemployment rate and year-on-year growth rate of the Harmonized Index of Consumer Prices (HICP) as a proxy for inflation. Because of inconsistencies in the reporting of inflation and unemployment series at the regional or provincial level, we stick with these measures at the national level. Observe, however, that GDP is not available at the regional level. Therefore, we use the regional real gross value added (GVA) as a proxy for the regional GDP. Figure 4 illustrates the average real gross value added and yearly growth rate, highlighting significant economic differences between regions in terms of growth and log levels.

Then we move to the four variables used to proxy the dynamics of the credit cycle. Public debt and cost-of-debt to GDP ratios are used to evaluate credit conditions at both the national and regional levels. In Figure 5, the top panel shows the variations in the cost-of-debt to GDP (on the left-hand side) and debt-to-GDP ratios (on the right-hand side) across regions. Generally, the central and southern regions have a higher level of debt relative to their GDP and also experience a higher cost of debt compared to the northern regions. We finally incorporate the spread between Italian and German 10-year constant maturity sovereign bond yields. To capture the effect of economic uncertainty we include the uncertainty index of Baker et al. (2016) for Italy.

The social environment can potentially affect the outcome of the credit collection process. For instance, higher levels of criminal activity, lower income, or less developed areas may directly affect the realized RR. To capture the socio-economic heterogeneity at the province level, we examine four variables: 1) the number of robberies, 2) intentional homicides, 3) unintentional homicides (all scaled per 100,000 inhabitants), and 4) the Life Quality Index published by the Italian financial newspaper *Il Sole 24 Ore*.⁶

The top panels of Figure 5 reveal that Milan, Rome and Naples are the provinces with the highest numbers of robberies and intentional and unintentional homicides per 100,000 inhabitants. The right panels of Figure 5 display instead the variation in the Life Quality Index across regions and provinces of Italy. The data highlights that regions and provinces in the north and the center tend to have higher average values of the Life Quality Index (538 and 521, respectively), compared to the south (419.95).

⁶This index consists of 90 indicators grouped into six macro-categories, namely wealth and consumption, business and work, environment and services, demographics and health, justice and security, and culture and leisure. Starting in 2019, each macro-category is made up of 15 sub-indicators, and provinces are awarded one thousand points for the best value and zero points for the worst in each sub-indicator. For each of the six macro-categories, a ranking is calculated based on the average score reported in the 15 sub-indicators, and the final ranking is based on the arithmetic mean of the six sector rankings. For more information visit <https://lab24.ilsole24ore.com/qualita-della-vita/>

Figure 4: Heterogeneity of Cost-of-debt and Debt to GDP ratios (top) and log-level of GDP and growth rate (bottom) at the region level. The regional GVA is used as a proxy for the GDP of the region.

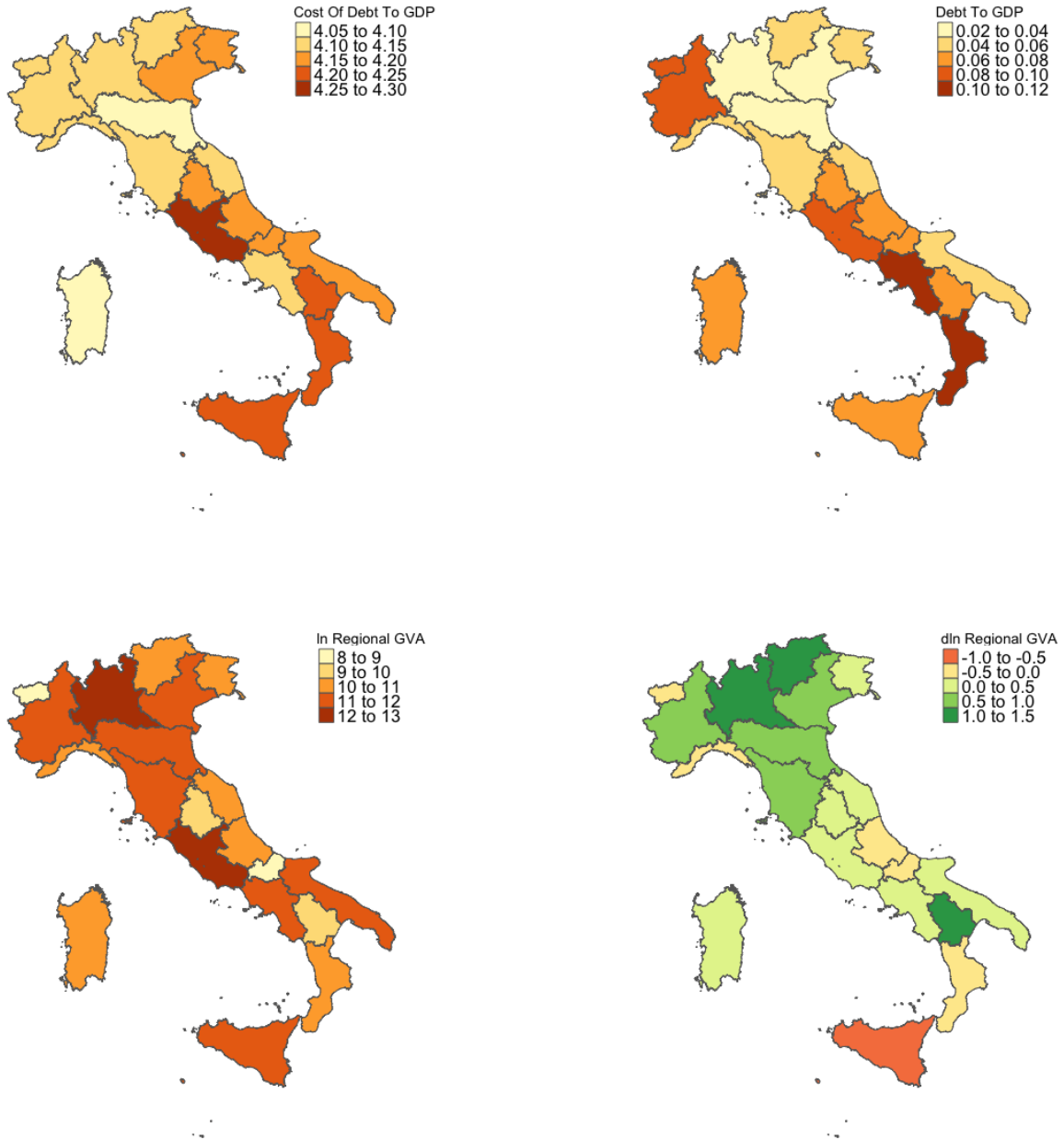
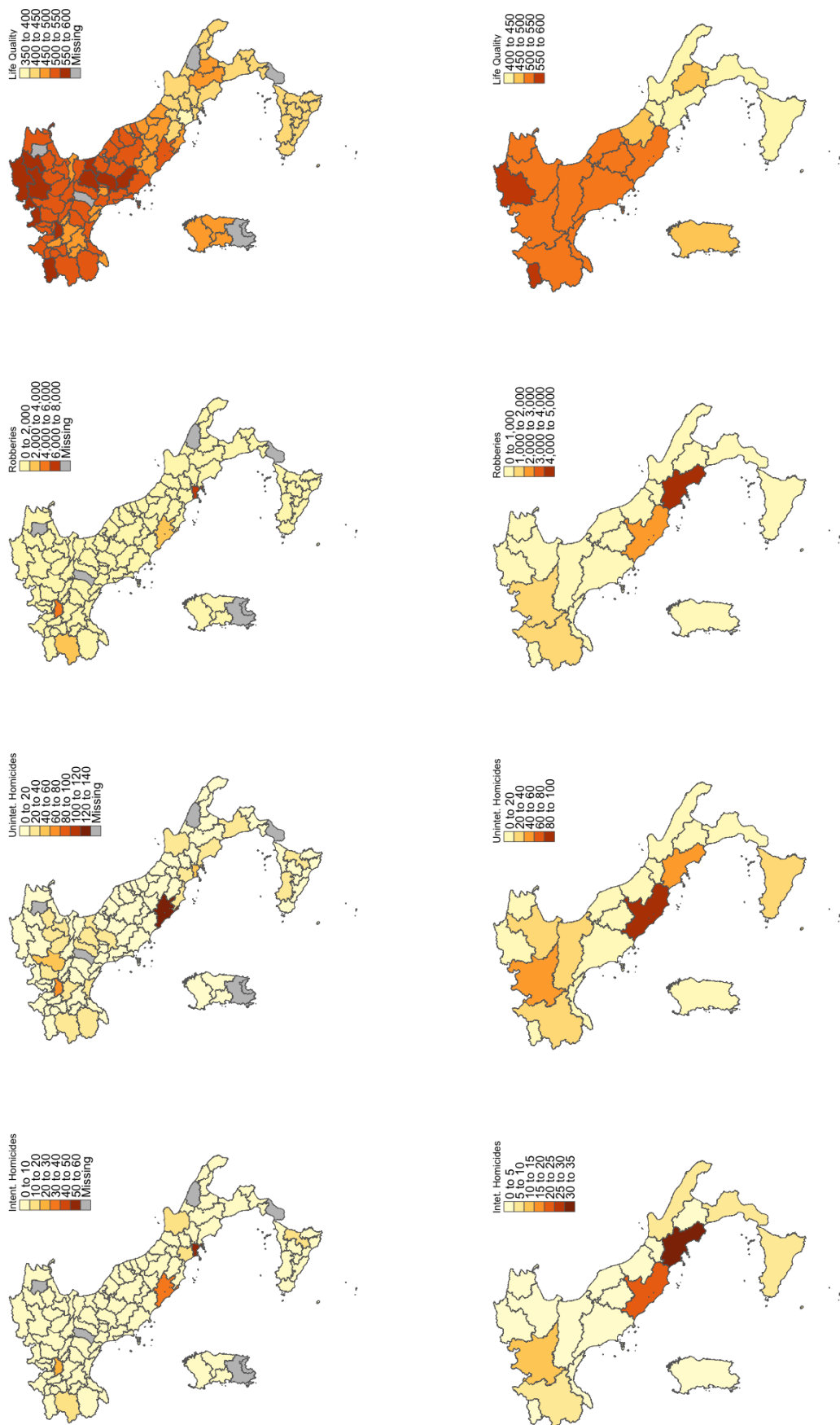


Figure 5: Heterogeneity of the number of intentional (left) and unintentional (center-left) homicides, robberies (center-right) and Life Quality Index (right) at province (top) and region (bottom) level.



Finally, we include fixed effect controls for the region of residence of the debtor and the industry from which the credit has originated (i.e., utilities, financing, banking, telecommunications, and commercials). As a result, we have a total of 43 predictive variables. We report the descriptive statistics of contract-specific and macroeconomic and social predictors in Tables 2 and 3.

Table 2: Descriptive statistics of the 13 considered macroeconomic and social (MS) variables. The last column indicates whether the frequency is monthly (M) or yearly (Y).

Variable	Abrev.	Mean	Std.Dev.	Min	Pctl. 25	Pctl. 75	Max	Freq.
Regional GVA Mln	ln GVA	119414.493	96610.849	4103.1	64286.2	141279	344629.2	Y
Regional GVA dln	dln GVA	0.108	1.975	-8.569	-0.258	1.149	8.947	Y
National Unemployment	Un	11.807	5.872	2.747	7.339	18.382	23.415	M
EPU Italy	EPU	110.972	32.992	31.702	89.521	130.025	241.018	M
HICP YoY	HCIP	0.861	0.825	-0.509	0.193	1.363	4.165	M
Regional Debt to GDP	R-DtGDP	0.07	0.029	0.024	0.042	0.09	0.149	Y
National Debt to GDP	N-DtGDP	129.787	9.979	103.9	132.5	134.8	135.4	Y
National Cost of Debt to GDP	N-CtGDP	4.148	0.484	3.6	3.8	4.6	5.2	Y
Spread ITA-GER 10Y bonds	Spread	171.549	65.719	21.066	130.701	205.274	460.318	M
Life Quality Index	LQ	490.15	66.603	359.97	425	545.55	641.49	Y
Intentional Homicides	INHom	11.03	16.52	0	1	13	109	Y
Unintentional Homicides	UNHom	30.931	31.788	0	10	40	138	Y
Robberies	Rob	1337.33	2202.379	4	91	1799	14045	Y

Table 4 indicates the sources for the MS indicators.

3.3 Portfolios of defaulted consumer credits

Third-party collectors typically acquire the right to recover pools of defaulted credit through auctions or direct negotiations, managing the recovery process on behalf of lenders and earning commissions based on the total recovered amount and collection actions performed. Therefore, studying recovery determinants at the portfolio level better reflects the business model of the third-party collection industry. These collectors often get some information about the sector, geographical area, and maximum duration for the collection process of the credit pool, but may lack detailed information about the characteristics of the individual loans packaged in the pool.

To mimic the day-to-day operations of debt collectors, we consider two scenarios. In the first scenario (scenario 1), third-party collectors have access to the precise information of each individual loan in the pool, allowing them to create forecasting models using constituent-level data. In this case, they can forecast the recovery rates of individual NPCs forming the pool, and therefore the forecasting performance can be evaluated both at individual credit and portfolio levels. In the second scenario (scenario 2), we assume that third-party collectors can access NPC information communicated in some aggregate way, at the portfolio level. Predictive models are thus trained based on a “portfolio-level predictor set”, and their output directly provides the forecast of the recovery rate for the entire pool. In this scenario, forecasting performance can be evaluated only at the portfolio level.

We now describe how we formed the portfolios of defaulted consumer credits for our analysis. We form a portfolio by randomly drawing without replacement samples of 100 individual credits that have been acquired during a specific quarter. We continue to form portfolios until we have less than 100 individual credits available for sampling in that quarter. We then move to the following quarter and repeat this operation for all the quarters in our samples, i.e. starting from 2007 Q1 up to 2019 Q4.

Table 3: Descriptive statistics of loan-specific variables having performed the stratified sampling procedure to Banking, Commercial, Financing, Telecommunications, and Utilities as described in Section 4.

Variable	Abrev.	Mean	Std.Dev.	Min	Pctl. 25	Pctl. 75	Max
Age	Age	55.335	17.572	21	43	67	100
Max Duration Collection	Max Dur.	90.877	76.28	0	30	120	365
Gender: F	F	0.272	0.445	0			1
Gender: M	M	0.425	0.494	0			1
Gender: Professionals	Prof.	0.006	0.076	0			1
Recovery Rate	RR	0.344	0.43	0	0	0.91	1
Credit to Recover	TtR	727.48	1104.19	50.01	185.13	742.672	10,000
Principal to Total to Recover	PTtR	0.881	0.186	0	0.874	1	1
Banking	Bank.	0.2	0.4	0			1
Commercial	Comm	0.2	0.4	0			1
Financing	Fin.	0.2	0.4	0			1
Telecommunications	Telco	0.2	0.4	0			1
Utilities	Util.	0.2	0.4	0			1
Region: Abruzzo	ABR	0.026	0.158	0			1
Region: Basilicata	BAS	0.008	0.091	0			1
Region: Calabria	CAL	0.034	0.181	0			1
Region: Campania	CAM	0.145	0.352	0			1
Region: Emilia Romagna	EMR	0.055	0.227	0			1
Region: Friuli Venezia Giulia	FVG	0.012	0.111	0			1
Region: Liguria	LIG	0.027	0.161	0			1
Region: Lombardia	LOM	0.135	0.342	0			1
Region: Marche	MAR	0.017	0.131	0			1
Region: Molise	MOL	0.006	0.076	0			1
Region: Piemonte	PIE	0.067	0.25	0			1
Region: Puglia	PIG	0.045	0.207	0			1
Region: Sardegna	SAR	0.011	0.102	0			1
Region: Sicily	SIC	0.107	0.309	0			1
Region: Trentino Alto Adige	TAA	0.006	0.076	0			1
Region: Tuscany	TUS	0.055	0.228	0			1
Region: Umbria	UMB	0.012	0.107	0			1
Region: Val d'Aosta	VDA	0.09	0.287	0			1
Region: Veneto	VEN	0.052	0.222	0			1

We consider portfolios of 100 credits because this offers an acceptable compromise between keeping enough observations for the training (and testing) set and maintaining the portfolio large enough to reduce the relevance of individual credits at the portfolio level. Figure 6 displays the number of portfolios over time.

In the portfolio, the total amount to recover and the recovered amount will simply be the sum of the respective amounts of the credits included in the portfolio. The recovery rate of the portfolio is defined as the weighted average

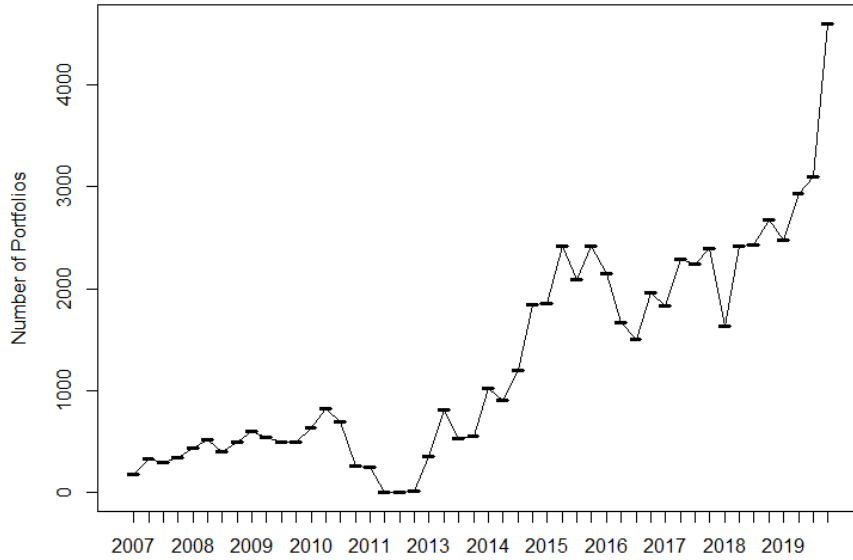
$$RR_{\pi} = \sum_i \pi_i RR_i \quad \text{where} \quad \pi_i = \frac{TtR_i}{\sum_j TtR_j}.$$

The “portfolio-level predictor set” for scenario 2 (contract-specific and MS predictors) coincide with the weighted averages of the corresponding predictors, where the weights are as above.

Table 4: List of sources for the considered macroeconomic and social (MS) variables. ISTAT stands for the Italian National Institute of Statistics (<https://www.istat.it/en/>) and BLB stands for Bloomberg.

Variable	Abrev.	Source
Regional GVA Mln	ln GVA	ISTAT
Regional GVA dln1	dln GVA	ISTAT
National Unemployment	Un	ISTAT
EPU Italy	EPU	www.policyuncertainty.com
HICP YoY	HCIP	ISTAT
Regional Debt to GDP	R-DtGDP	ISTAT
National Debt to GDP	N-DtGDP	ISTAT
National Cost of Debt to GDP	N-CtGDP	ISTAT
Spread ITA-GER 10Y bonds	Spread	BLB
Life Quality Index	LQ	https://lab24.ilsole24ore.com/qualita-della-vita/
Intentional Homicides	INHom	ISTAT
Unintentional Homicides	UNHom	ISTAT
Robberies	Rob	ISTAT

Figure 6: Number of portfolios over time.



4 Methodology

When third-party collectors have precise information on the individual loans in the pool, they can leverage constituent-level data to estimate the forecasting models. Following the methodology of previous studies in the recovery rate forecasting literature (Nazemi et al., 2017, 2018; Nazemi and Fabozzi, 2018; Gambetti et al., 2022), we use stratified sampling to create training and testing sets with the same proportions of NPCs originating from different industries (telecommunication, utilities, banking, commercial loans, or consumer finance). To keep the computational cost of the study under control, but, more importantly, to save enough observations in the test set for evaluating the forecasting performance at the portfolio level when using models trained using constituent-level data, we consider a 10-90% split between training and

testing sets. This resulted in 124,193 individual NPCs for training and 1,080,562 individual NPCs for testing equally spread across industries. When evaluating forecasting performance at the portfolio level, we aggregate the 1,080,562 individual NPCs in the test set into 10,800 portfolios, as detailed in Section 3.3.⁷

When third-party collectors can access information only at the portfolio level, we estimate the forecasting models and evaluate the performance using portfolio-level data only. In this case, we use the 6,493,794 observations to build 62,082 portfolios as indicated in Section 3.3 and consider a standard 70% - 30% split between training and testing sets.

As a robustness check, we design a prediction exercise in which the third-party collector ensures that the training set only features portfolios preceding the ones included in the testing set. This approach (also known as “out-of-time”) controls for potential structural breaks in the data-generating process, which could potentially lead to a look-ahead bias in the predictions (Kalotay and Altman, 2016; Nazemi et al., 2022). In this exercise, the third-party collector lacks data on the specific loans forming the portfolios and uses portfolios of NPCs acquired before January 2017 for training, while credits acquired after January 2017 are left for testing. This corresponds to an approximately 50% - 50% split between training and testing sets permitting us to have enough observations to train data-intensive predictive models while keeping enough variability of yearly MS series.⁸

5 Forecasting methods and evaluation criteria

Several papers have highlighted the potential of machine learning (ML) techniques in forecasting recovery rates (Qi and Zhao, 2011; Loterman et al., 2012; Yao et al., 2017; Nazemi et al., 2018; Nazemi and Fabozzi, 2018; Hurlin et al., 2018; Nazemi et al., 2017; Bellotti et al., 2021; Nazemi et al., 2022; Gambetti et al., 2022), overcoming many of the limitations of the traditional multivariate linear and beta regressions used in earlier studies. For example, because recovery rates are defined in the $[0, 1]$ interval, predictions arising from a linear regression framework may lead to values outside that interval. Also, the observed bi-modal distribution of individual credit recovery rates contrasts with the assumption of uni-modality underlying the beta regressions. These facts limit the applicability of linear and beta regressions in this context.

In this study, we employ seven well-known predictive models in credit scoring and recovery rate modeling literature (Loterman et al., 2012; Bellotti et al., 2021). We include gradient-boosted regression trees, random forests, (bagged and individual) multivariate adaptive splines. We also keep the linear regression model estimated with ordinary least squares (OLS) and beta regression (Ferrari and Cribari-Neto, 2004) as benchmarks. Among these models, gradient-boosted regression trees, in particular, have proven to offer solid forecasting performance while remaining computationally tractable (Schmitt, 2022). We also explore the possibility of linear and nonlinear model combinations, which may offer diversification gains and are attractive when we cannot identify *ex-ante* the best single model or the employed predictive schemes cover a wide spectrum of modeling assumptions (Atiya, 2020; Roccazzella et al., 2022).

⁷Stratified sampling offsets the bias toward sectors that count more NPCs at the cost of reducing the actual observations. The large difference in the number of NPCs originating from the telecommunication and utility sectors compared to the banking and finance (see Table 1) motivates why we perform the stratified sampling at the sector level instead of, for instance, gender.

⁸We also performed this analysis using the NPCs acquired before January 2016 were used for training, while credits acquired after January 2016 were used for testing. Results are robust to this specification. The latter are available upon request.

Table 5 reports the predictive algorithms and the forecast combinations employed in this study. We refer to the Supplementary Material A for a detailed description of the predictive algorithms employed, and we now proceed to describe the three forecast combinations (FC) methods briefly.

Table 5: List of considered algorithms and corresponding R packages.

Description	Acronym	R package	Reference
Equally weighted average forecast	EW FC		
Optimal FC(+)	Opt+ FC	quadprog	
Averaged neural networks	ANN FC	avnnet	Ripley (1996)
OLS Linear regression	LM	lm	
Beta regression	beta	betareg	Ferrari and Cribari-Neto (2004)
Multivariate adaptive regr. splines	MARS	earth	Friedman (1991)
Bagged MARS	B-MARS	bagEarth	Friedman (1991)
Random Forest	RF	ranger	Breiman (2001)
Gradient Boosted Regr. Trees	GBRT	gbm	Friedman (2002)
XGB Trees	XGBt	xgboost	Chen and Guestrin (2016)

Notes. The predictive algorithms can be retrieved via the R library **caret** (Kuhn, 2008) and refer to Kuhn and Johnson (2013) for a textbook treatment of the R packages. Hyper-parameters are tuned via 5-fold cross-validation. We refer to the Appendix for R pseudo-code.

We consider a fixed linear weighting scheme in the simple average forecast (EW FC), and two optimization-based combination methods that are estimated from the data: a) the minimum-variance linear combination of forecast with nonnegative weights (Opt+ FC) and, b) the neural-network based nonlinear combination of forecasts (ANN FC).

Assuming that the individual forecasts are unbiased⁹, Opt+ FC minimizes the variance of the aggregate forecast error. In this case, restricting the weights to sum to one keeps the aggregated forecast unbiased as well (Timmermann, 2006). Following Granger and Ramanathan (1984), with the sum-to-one and non-negativity constraints on the combining weights, the optimization problem becomes

$$\mathbf{w}^* := \underset{\mathbf{w} \in \mathcal{W}}{\operatorname{argmin}} \mathbf{w}^T \hat{\Sigma} \mathbf{w} \quad \text{where} \quad \mathcal{W} := \{\mathbf{w} \in \mathbb{R}^7 \mid \mathbf{1}^T \mathbf{w} = 1, \min \mathbf{w} \geq 0\}, \quad (1)$$

where $\hat{\Sigma}$ is the 7×7 -sample covariance matrix of the models' prediction errors.

This problem can be solved efficiently using quadratic programming or imposing the Karush-Kuhn-Tucker conditions. Restricting the weights to be non-negative and to sum to one implicitly selects which forecasts to combine and reduces the estimation error by inducing a shrinkage-like effect on the covariance matrix of forecast errors (Jagannathan and Ma, 2003).¹⁰

With ANN FC, we explore the relevance of nonlinear weighting schemes for the forecast combination. Specifically, we opt for averaged neural networks that take the forecasts produced by RF, GBRT, XGBt, MARS, B-MARS, Beta, and OLS linear regression models as inputs and aggregate these into one prediction. ANN FC is an ensemble of one-hidden-layer feed-forward neural networks (Ripley, 1996) and with

⁹This assumption holds in-sample as all forecasting models include an intercept. Remark that, under this assumption, minimizing the variance of the aggregate forecast error coincides with minimizing the mean square error of the aggregate forecast error.

¹⁰As a robustness check, we also considered the optimal and robust combinations of forecasts proposed by Roccazzella et al. (2022). Results are comparable and available upon request.

the *sigmoid* activation function. The ensemble counts 20 neural networks that are initialized by using different starting values for the parameters to estimate and that include a *decay* factor to penalize large coefficients.¹¹ These features moderate the risk of over-fitting.

Opt+ FC and ANN FC combining weights are estimated in a *combination fold*, i.e., a sub-sample of the training set that is not used to estimate the RF, GBRT, XGBt, MARS, B-MARS, Beta, and OLS linear regression models. The size of the combination fold is 3% of the training set. This offers an acceptable compromise between having enough observations to train the data-intensive ANN FC and losing information for the estimation of the other predictive algorithms that, as a result, use 97% of the training set. In the case of machine learning methods, hyperparameters are tuned using 5-fold cross-validation.

We consider the widely used evaluation criteria for forecasting models: the root mean square error (RMSE), the mean absolute error (MAE), and the R^2 , defined as

$$R_j^2 = 1 - \frac{\sum_i^{n_{test}} (y_i - \hat{y}_{i,j})^2}{\sum_i^{n_{test}} (y_i - \bar{y})^2},$$

where n_{test} is the number of observations in the test set, y_i is the i^{th} realization of the recovery rate, $\hat{y}_{i,j}$ is the i^{th} forecast of the recovery rate obtained following the strategy j and $\bar{y}$ is the average recovery rate in the training set. Notice that R^2 can be negative, which signals that the forecast method in question performs worse than a strategy considering the average recovery rate in the training set as a forecast.

We compute the averages of out-of-sample performance measures for the models via bootstrap, considering samples of 100 observations and repeating the procedure 10,000 times. As in Nazemi et al. (2022), we use the Mann–Whitney test to assess whether the differences in performance when including MS indicators predictors are statistically significant.

Finally, we also consider the model confidence set (hereafter MCS, Hansen et al., 2011), which tests whether a subset of methods enters jointly in the superior set of models by repeatedly testing the null hypothesis of equal predictive performance for a given significance level α .¹²

6 Results

We first look at the performance of the out-of-sample forecasting exercise on individual NPCs and portfolios of NPCs. We then employ Accumulated Local Effects (ALE) plots (Apley and Zhu, 2020) to shed light on how individual explanatory variables impact the model predictions. This enables us to determine whether our predictive models also capture theoretically sound relationships between recovery rates and conditioning variables.

¹¹Coefficients (called weights in neural networks) are prevented from becoming too large by penalizing their L2 norm similarly to the L2 penalty in the Ridge regression. The intensity of the penalization, called weight decay in neural networks, and the number of neurons are tuned via 5-fold cross-validation. We set the number of neural networks in the ensemble to 20 to limit the computational burden. Results remain comparable when also considering larger ensembles of 30 and 40 networks.

¹²For the MCS, we consider the quadratic loss function and compute the p -values via bootstrapping (5,000 replications).

6.1 Forecasting performance

Table 6 reports the models’ forecasting performance using various metrics (RMSE, MAE and R^2) with (columns 2 to 4) and without (columns 5 to 7) MS predictors as well as the performance gains (last three columns) in four different specifications: at the individual loan level (panel a), at the portfolio level using individual-level predictors (scenario 1, panel b), at the portfolio level using portfolio-level predictors (scenario 2, panel c) and at the portfolio level using portfolio-level predictors but in an “out-of-time” setup (time-consistent scenario 2, panel d). For each panel, we report the list of models joining the 5% model confidence set. The best performance measures for each specification are highlighted in bold.

Let us first analyze the difference in performance when looking at the individual *vs.* portfolio level. Comparing panels a) and b), we observe a substantial enhancement in predictive accuracy when working with portfolios. For instance, if we choose Opt+ FC when MS predictors are excluded, the R^2 value rises from 28.27% for individual loans to 51.66 % in the context of portfolios. This suggests that the idiosyncratic risk of individual credits’ recovery rates tends to be diversified away. This is illustrated in Figure 7, which shows how the aggregation into portfolios impact the RR distribution: it becomes uni-modal and its variance shrinks. This is consistent with the theoretical result on the asymptotic distribution of the averaged RR of an equally weighted infinitely granular pool, where defaults are controlled by a Gaussian copula (see the discussion in the Supplementary Material C). Comparing panels b) and c), we study the benefit of possessing a constituents-level information set when forecasting aggregate RRs. The forecasting performance significantly deteriorates in absence of detailed information on the loans in the portfolio. For instance, if we again choose Opt+ FC when MS predictors are excluded, the R^2 value decreases from 51.66% to 32.40% when one only possess portfolio-level information. In addition, Table 6 suggests that forecast combination techniques provide the most effective approach for predicting RRs. In particular, the use of artificial neural networks as a nonlinear combination strategy (ANN FC) is encouraging since it is consistently included in the superior group of models and attains the lowest MAE and RMSE values.

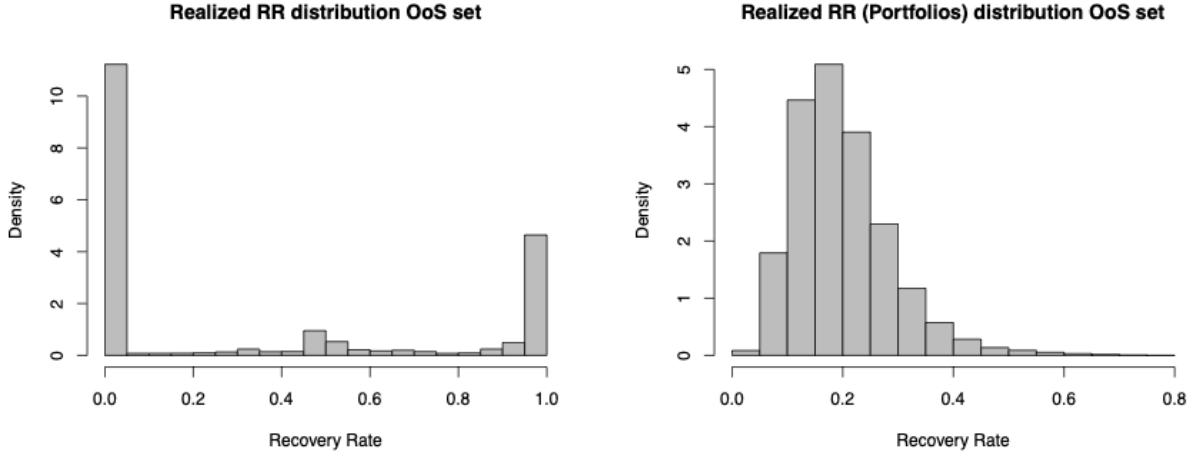
Let us now study the impact of MS indicators in the performance. The last three columns of Table 6 display the percentage differences in performance measures when MS indicators are excluded from the predictor set. Performance significantly deteriorates: this is evident when forecasting the recovery rates of defaulted individual credits (panel a), but it is even more apparent at the portfolio level (panels b and c). When examining individual loans, the maximum improvement in R^2 is approximately 3 percentage points (XGBt in panel a). However, the improvement in portfolios is much more pronounced. For instance, including MS predictors in RF helps increasing the R^2 by more than 6 percentage points (RF in panel c).

Finally, we control for potential look-ahead bias in the models adopting an out-of-time setup. For conciseness, we report the performance associated with the portfolio level in scenario 2. Panel d) reports the performance for forecasting models trained exclusively on portfolio-level information and NPCs obtained before January 2017. Compared to the random training and testing setup in panel c), we note that performance generally worsens. For instance, the R^2 value drops from approximately 37% to 28% for the ANN FC. Yet, the positive impact of incorporating MS indicators remains evident. Specifically, nine out of ten models improve their RMSE and R^2 , and eight out of ten improve the MAE. The size of this improvement is particularly noteworthy. For instance, the ANN FC model sees a rise of over 7 percentage points in its R^2 , while the Beta and OLS exhibit an increase of more than 15 p.p.

Table 6: Performance measures for each specification; best scores are shown in bold. The last 3 columns indicate the loss differentials when excluding MS and the respective significance levels for the Mann-Whitney test with **, *, and **, for the 1%, 5%, and 10% levels. We report the superior set of models at the 5% significance level and the L2 loss denoting whether the model include the MS or not with the acronyms *MS* or *no MS*.

Predictors set excluding MS				Predictors set including MS		RMSE		% Difference		MAE		R2	
RMSE	MAE	R2		RMSE	MAE	R2		RMSE		MAE		R2	
a) Out of Sample - Individual Loans													
MARS	.3840	.3213	.1980	.3814	.3176	.2089		0.69%		1.17%		-1.08	
GBRT	.3676	.3004	.2646	.3632	.2952	.2818		1.20%		1.76%	*	-1.72	*
XGBt	.3736	.3040	.2396	.3661	.2868	.2695		2.06%	**	6.00%	**	-2.99	**
RF	.3673	.2915	.2654	.3615	.2867	.2880		1.58%		1.67%		-2.26	*
B-MARS	.3847	.3249	.1952	.3822	.3210	.2054		0.65%		1.21%		-1.02	
Beta	.4282	.3990	.0046	.4247	.3946	.0205		0.82%	*	1.13%	**	-1.59	**
OLS	.4027	.3473	.1189	.3987	.3422	.1364		1.01%	*	1.49%	**	-1.74	**
EW FC	.3734	.3241	.2422	.3684	.3180	.2624		1.37%	*	1.92%	**	-2.03	**
Opt+ FC	.3630	.2960	.2827	.3567	.2884	.3074		1.78%	*	2.63%	**	-2.47	**
ANN FC	.3633	.2935	.2814	.3561	.2801	.3093		2.03%	*	4.80%	**	-2.79	**
MCS: {ANN FC (MS)}													
b) Out of Sample - Portfolios with Individual Level Predictors (scenario 1)													
MARS	.0693	.0548	.4199	.0687	.0547	.4285		0.89%		0.20%		-0.86	
GBRT	.0635	.0503	.5139	.0610	.0485	.5501		4.06%	**	3.64%	**	-3.63	**
XGBt	.0718	.0588	.3763	.0659	.0534	.4747		9.01%	**	10.27%	**	-9.84	**
RF	.0642	.0514	.5023	.0632	.0510	.5170		1.59%		0.74%		-1.47	
B-MARS	.0689	.0547	.4287	.0663	.0528	.4693		3.98%	**	3.57%	**	-4.06	**
Beta	.1093	.0899	-.440	.1012	.0846	-.237		8.01%	**	6.27%	**	-20.28	**
OLS	.0796	.0637	.2385	.0737	.0590	.3429		7.96%	**	7.95%	**	-10.44	**
EW FC	.0681	.0552	.4416	.0648	.0526	.4937		5.20%	**	4.98%	**	-5.21	**
Opt+ FC	.0633	.0507	.5166	.0613	.0494	.5459		3.25%	**	2.70%	**	-2.92	**
ANN FC	.0626	.0497	.5268	.0599	.0475	.5656		4.48%	**	4.54%	**	-3.88	**
MCS: {ANN FC (MS)}													
c) Out of Sample - Portfolios with Portfolio Level Predictors (scenario 2)													
MARS	.0745	.0556	.2820	.0725	.0540	.3198		2.76%	**	2.96%	**	-3.78	**
GBRT	.0733	.0548	.3050	.0705	.0528	.3570		3.97%	**	3.79%	**	-5.20	**
XGBt	.0740	.0559	.2912	.0718	.0542	.3314		3.06%	**	3.14%	**	-4.02	**
RF	.0734	.0553	.3026	.0700	.0529	.3652		4.86%	**	4.54%	**	-6.26	**
B-MARS	.0738	.0552	.2974	.0736	.0548	.3022		0.27%		0.73%		-0.48	
Beta	.0790	.0583	.2013	.0768	.0567	.2438		2.86%	*	2.82%	*	-4.25	**
OLS	.0790	.0586	.1998	.0768	.0571	.2421		2.86%	*	2.63%	*	-4.23	**
EW FC	.0733	.0551	.3078	.0711	.0535	.3489		3.09%	*	2.99%	**	-4.11	**
Opt+ FC	.0723	.0545	.3240	.0697	.0526	.3715		3.73%	*	3.61%	**	-4.75	**
ANN FC	.0724	.0546	.3205	.0697	.0524	.3690		3.87%	*	4.20%	**	-4.85	**
MCS: {Opt+ FC (MS); ANN FC (MS)}													
d) Out of Time - Portfolios with Portfolio Level Predictors (scenario 2)													
MARS	.0813	.0592	.1590	.0772	.0589	.2414		5.31%	**	0.54%		-8.24	**
GBRT	.0798	.0584	.1919	.0760	.0562	.2687		5.05%	**	3.93%	**	-7.68	**
XGBt	.0852	.0636	.0802	.0809	.0607	.1712		5.33%	**	4.77%	**	-9.09	**
RF	.0780	.0576	.2280	.0785	.0617	.2153		-0.60%	**	-6.66%	**	1.27	**
B-MARS	.0820	.0601	.1454	.0767	.0577	.2524		6.98%	**	4.24%	**	-10.70	**
Beta	.0848	.0632	.0891	.0774	.0582	.2396		9.52%	**	8.65%	**	-15.05	**
OLS	.0868	.0656	.0434	.0788	.0598	.2119		10.25%	**	9.78%	**	-16.86	**
EW FC	.0811	.0601	.1665	.0758	.0572	.2721		7.04%	*	5.19%	**	-10.56	**
Opt+ FC	.0784	.0578	.2199	.0764	.0590	.2580		2.62%	**	-2.05%	**	-3.80	**
ANN FC	.0793	.0584	.2015	.0753	.0572	.2796		5.36%	**	2.03%	**	-7.81	**
MCS: {ANN FC (MS)}													

Figure 7: Distributions of the realized recovery rate for defaulted individual credits (left) and portfolios of consumer credits (right).



6.2 Visualizing the Accumulated Local Effects

To examine the average effect of each predictor, we use ALE plots which rely on a localization to analyze model predictions.¹³ The vertical axes measure the differences for the average prediction, while the horizontal axes the domain of the corresponding variable.

We first analyze whether the industry originating the credit affects average recovery rates. ALE plots in Figure 8 confirm what we have already observed in Table 1. Setting the *Utility* sector as a baseline, we see that models predict greater recovery rates when the proportion of credits from the banking sector increases (see *Banking*). We observe a similar but less accentuated effect related to quota of credits from the financing industry (see *Financing*) and commercial credits (see *Commercial*); conversely, the greater the proportion of defaulted telecommunication credits in the portfolio, the lower the expected recovery rate. This highlights that findings about NPC RRs related to a specific sector cannot be easily generalized to other industries, in contrast with what previous research suggested (Nazemi et al., 2022). For example, the greater expected recovery rate associated with the *Banking* quota may reflect a better screening ability or be consistent with a different debtor behavior (e.g., the fear of being refused future loans from banks).

Figure 9 illustrates the impact of some loan-specific variables. A greater proportion of interest and ancillary fees in the total amount to recover leads to a lower expected recovery rate (see *Principal to Total to Recover*). We confirm the negative correlation between the total amount to recover and the recovery rate (see *Credit to Recover*) previously observed in other papers (Bellotti et al., 2021; Nazemi et al., 2022). Interestingly, being granted more time to perform the collection does not seem to enhance expected recovery rates, on the contrary (see *Max Time to Recover*).

¹³Another visualization technique is Partial Dependence (PD) plots, but they tend to be more computationally expensive and may produce inaccurate results when predictors are highly correlated. For a more in-depth discussion of ALE, PD, and marginal plots, we refer readers to Apley and Zhu (2020).

Figure 8: ALE plots: impact of the sector.

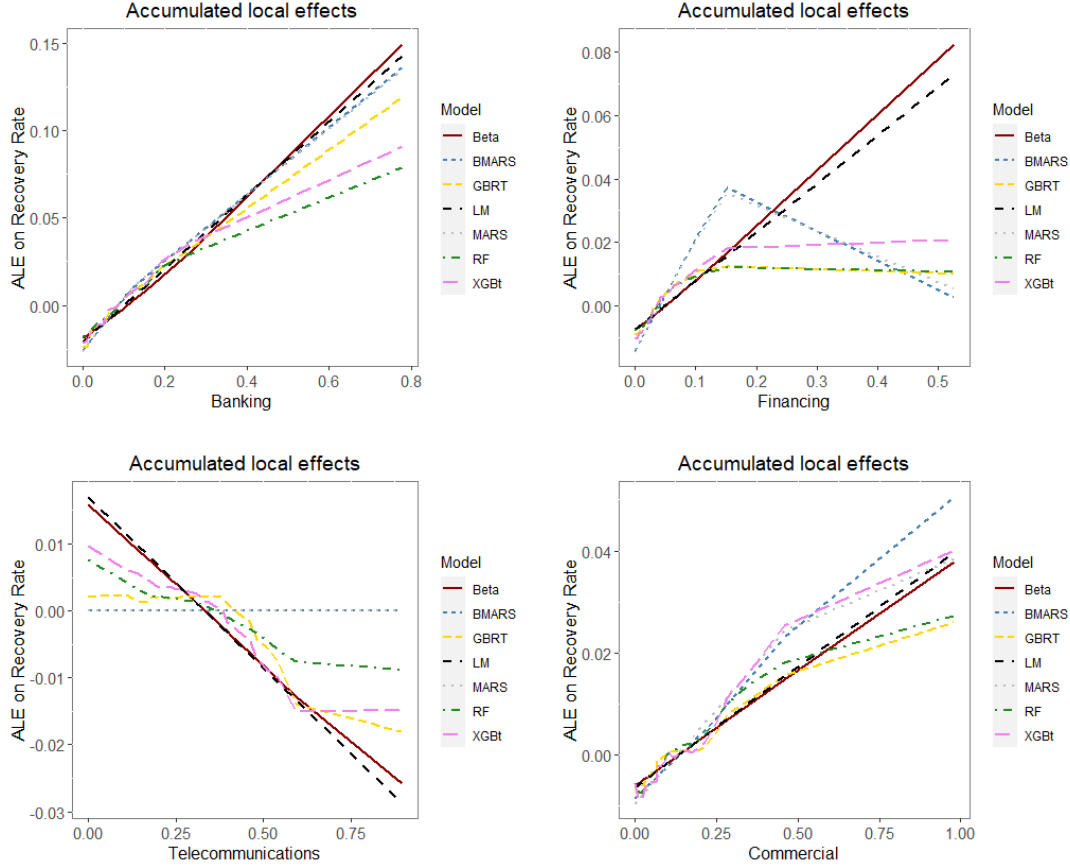
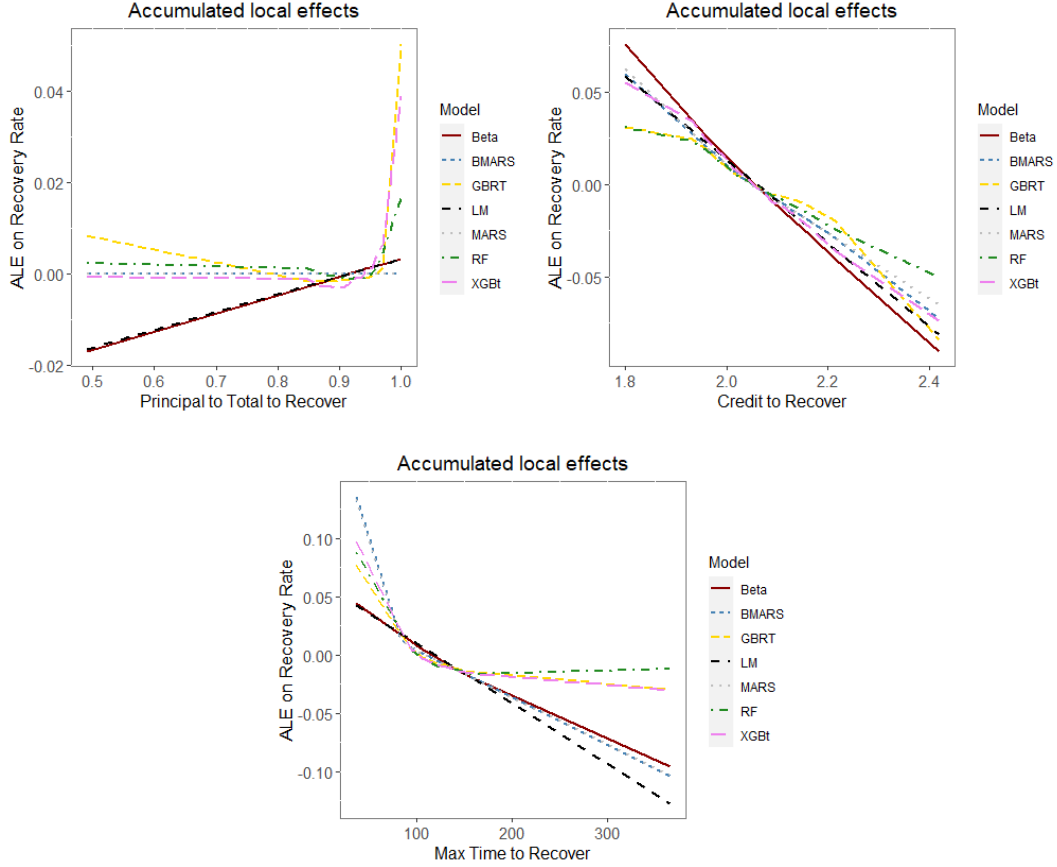


Figure 11 focuses on macroeconomic indicators. We notice that higher levels of public debt interest are associated with lower expected recovery rates (see *Cost of Debt to GDP*). In addition, low uncertainty seems to be associated with higher expected RRs (see *EPU* in Figure 11). Among the MS variables, the *EPU* has the largest effect, extending the findings of Gambetti et al. (2019) to defaulted consumer credits.

National inflation seems positively associated with higher expected RRs, and the estimated ALE varies across models (see *HICP YoY*). However, it is worth noting that within our sample, the inflation level never went above 4% and we cannot properly assess the effect of high inflation rates. Finally, a high level of real regional output is also associated with a greater expected recovery rate. The relationship is highly nonlinear, and the size of the effect is modest and only captured by gradient-boosted trees (see *Regional GVA*).

Figure 9: ALE plots: impact of some loan-specific variables.



7 Conclusions

Forecasting recovery rates is crucial for all stakeholders involved in estimating the loss of non-performing consumer credits. Our findings demonstrate the relevance and importance of macroeconomic variables and social indicators for the performance of predictive models, complementing the existing literature.

Our results hold irrespective of whether we aim at forecasting the recovery rate for an individual loan or a portfolio of loans, and are relatively stronger in the latter case. This result might be useful to industry participants as it is closer to their day-to-day business operations. Moreover, the analysis of accumulated local effects shows that the relationship between expected recovery rates and predictors related to the dynamics of the business cycle is intuitive: a deterioration of the credit conditions, signaled by lower real activity, increasing cost-to-GDP ratio, and increasing economic uncertainty, is associated to lower expected recovery rates.

Finally, producing a combination of models seems to be highly beneficial, confirming the intuition that it helps to diversify model risk. The improvement is particularly important for non-linear forecast combination schemes relying on neural networks.

Acknowledgments

The work of Frédéric Vrins is supported by the Fonds de la Recherche Scientifique - (F.S.R.-FNRS) grant CDR J.0037.18 and by the grant ARC 18-23/089 of the Belgian Federal Science Policy Office.

Figure 10: ALE plots: impact of some regional (Sicily and Val d'Aosta) and criminality indicators.

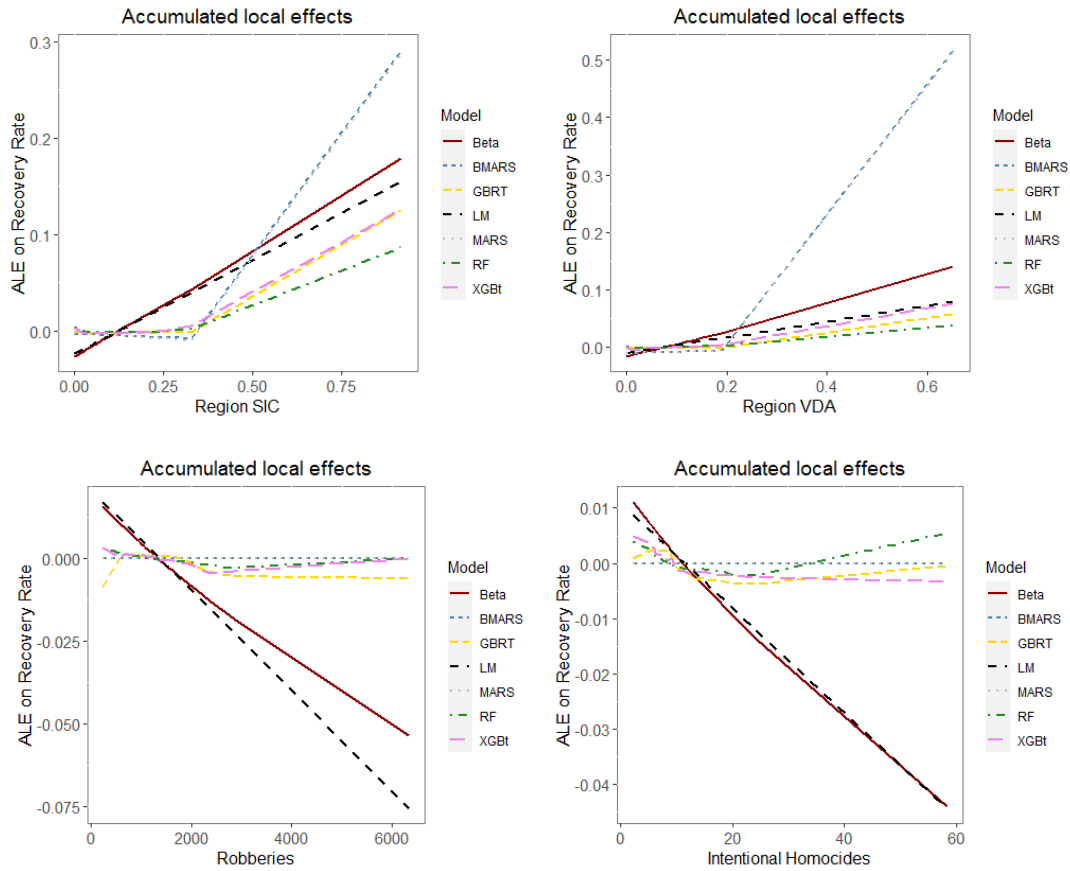
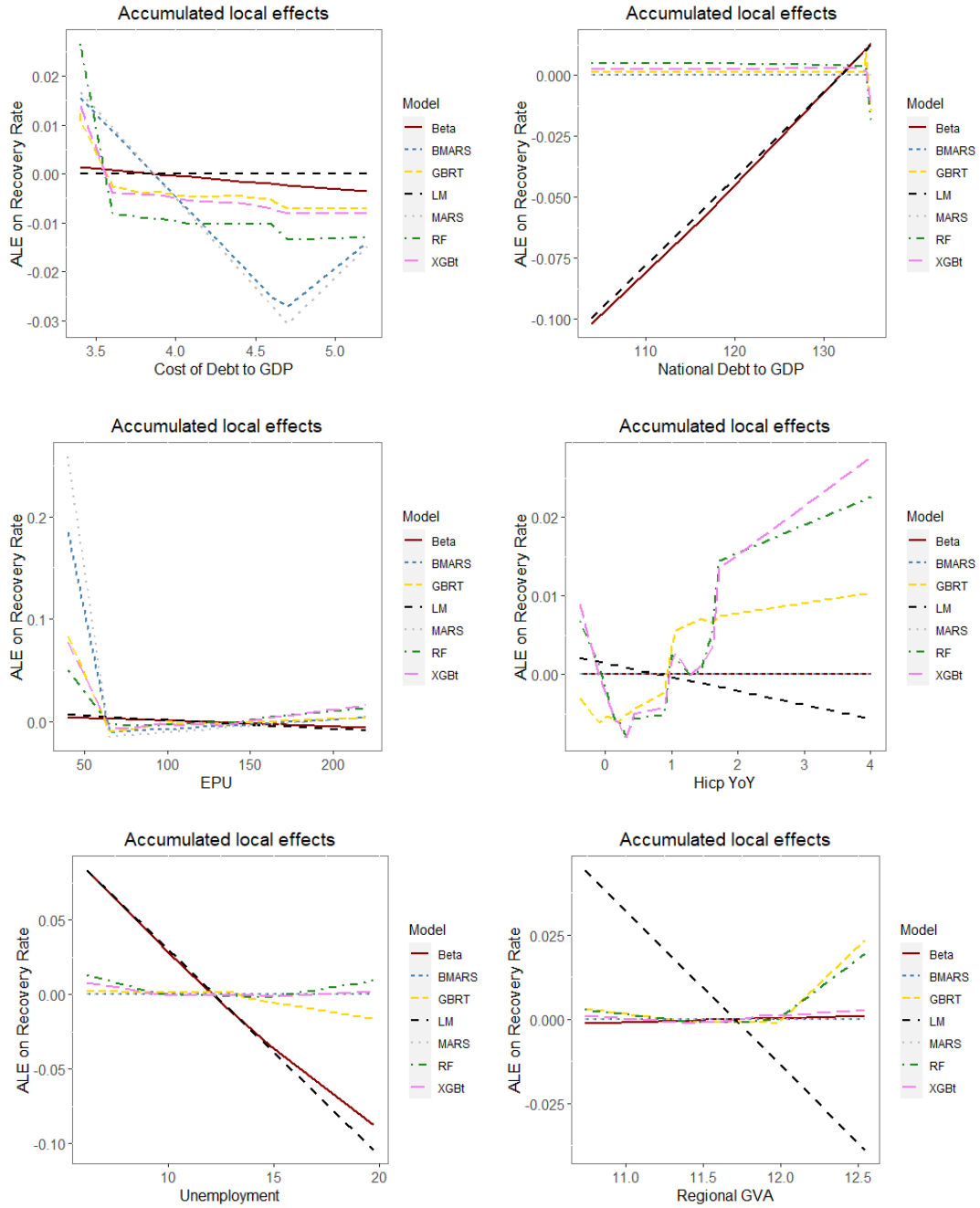


Figure 11: Accumulate Local Effect Plots.



References

- Altman, E., B. Brady, A. Resti, and A. Sironi (2005). The link between default and recovery rates: theory, empirical evidence, and implications. *The Journal of Business* 78(6), 2203–2227.
- Apley, D. W. and J. Zhu (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society Series B* 82(4), 1059–1086.
- Atiya, A. F. (2020). Why does forecast combination work so well? *International Journal of Forecasting* 36(1), 197–200.
- Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics* 131(4), 1593–1636.
- Beck, T., J. Grunert, W. Neus, and A. Walter (2017). What determines collection rates of debt collection agencies? *Financial Review* 52(2), 259–279.
- Bellotti, A., D. Brigo, P. Gambetti, and F. Vrms (2021). Forecasting recovery rates on non-performing loans with machine learning. *International Journal of Forecasting* 37(1), 428–444.
- Bellotti, T. and J. Crook (2012). Loss given default models incorporating macroeconomic variables for credit cards. *International Journal of Forecasting* 28(1), 171–182.
- Breiman, L. (1996). Bagging predictors. *Machine Learning* 24(2), 123–140.
- Breiman, L. (2001). Random forests. *Machine Learning* 45(1), 5–32.
- Chen, T. and C. Guestrin (2016). XGBoost. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM.
- Dermine, J. and C. N. de Carvalho (2006). Bank loan losses-given-default: A case study. *Journal of Banking & Finance* 30(4), 1219–1243.
- European Central Bank (2016). Bank lending survey for the euro area: glossary. Technical report.
- Fedaseyeu, V. (2020). Debt collection agencies and the supply of consumer credit. *Journal of Financial Economics* 138(1), 193–221.
- Ferrari, S. and F. Cribari-Neto (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics* 31(7), 799–815.
- Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics* 19(1), 1–67.
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Analysis* 38(4), 367–378.
- Gambetti, P., G. Gauthier, and F. Vrms (2019). Recovery rates: Uncertainty certainly matters. *Journal of Banking & Finance* 106, 371–383.
- Gambetti, P., F. Roccazzella, and F. Vrms (2022). Meta-learning approaches for recovery rate prediction. *Risks* 10(6).

- Gordy, M. B. (2003). A risk-factor model foundation for ratings-based bank capital rules. *Journal of Financial Intermediation* 12(3), 199–232.
- Granger, C. W. J. and R. Ramanathan (1984). Improved methods of combining forecasts. *Journal of Forecasting* 3(2), 197–204.
- Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica* 79(2), 453–497.
- Hurlin, C., J. Leymarie, and A. Patin (2018). Loss functions for loss given default model comparison. *European Journal of Operational Research* 268(1), 348–360.
- Jagannathan, R. and T. Ma (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance* 58(4), 1651–1683.
- Kalotay, E. A. and E. I. Altman (2016). Intertemporal forecasts of defaulted bond recoveries and portfolio losses. *Review of Finance* 21(1), 433–463.
- Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software, Articles* 28(5), 1–26.
- Kuhn, M. and K. Johnson (2013). *Applied Predictive Modeling*. New York. Springer.
- Leow, M., C. Mues, and L. Thomas (2014). The economy and loss given default: evidence from two UK retail lending data sets. *Journal of the Operational Research Society* 65(3), 363–375.
- Loterman, G., I. Brown, D. Martens, C. Mues, and B. Baesens (2012). Benchmarking regression algorithms for loss given default modeling. *International Journal of Forecasting* 28(1), 161–170.
- Nazemi, A., F. Baumann, and F. J. Fabozzi (2022). Intertemporal defaulted bond recoveries prediction via machine learning. *European Journal of Operational Research* 297(3), 1162–1177.
- Nazemi, A. and F. J. Fabozzi (2018). Macroeconomic variable selection for creditor recovery rates. *Journal of Banking & Finance* 89, 14–25.
- Nazemi, A., K. Heidenreich, and F. J. Fabozzi (2018). Improving corporate bond recovery rate prediction using multi-factor support vector regressions. *European Journal of Operational Research* 271(2), 664–675.
- Nazemi, A., F. F. Pour, K. Heidenreich, and F. J. Fabozzi (2017). Fuzzy decision fusion approach for loss-given-default modeling. *European Journal of Operational Research* 262(2), 780–791.
- Nazemi, A., H. Rezazadeh, F. J. Fabozzi, and M. Höchstötter (2022). Deep learning for modeling the collection rate for third-party buyers. *International Journal of Forecasting* 38(1), 240–252.
- Probst, P. and A.-L. Boulesteix (2017). To tune or not to tune the number of trees in random forest. *Journal of Machine Learning Research* 18(1), 6673–6690.
- Qi, M. and X. Zhao (2011). Comparison of modeling methods for loss given default. *Journal of Banking & Finance* 35(11), 2842–2855.

- Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press.
- Roccazzella, F., P. Gambetti, and F. Vrms (2022). Optimal and robust combination of forecasts via constrained optimization and shrinkage. *International Journal of Forecasting* 38(1), 97–116.
- Schmitt, M. (2022). Deep learning vs. gradient boosting: Benchmarking state-of-the-art machine learning algorithms for credit scoring.
- Smithson, M. and J. Verkuilen (2006). A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods* 11, 54–71.
- Timmermann, A. (2006). Chapter 4 forecast combinations. In *Handbook of Economic Forecasting*, pp. 135–196. Elsevier.
- Yao, X., J. Crook, and G. Andreeva (2017). Enhancing two-stage modelling methodology for loss given default with support vector machines. *European Journal of Operational Research* 263(2), 679–689.

A Forecasting Models

A.1 Linear and Beta regression models

Let p be the number of predictive variables and n be the sample size. We study the relationship between the n -dimensional vector of recovery rates $\mathbf{y}$ and the $n \times p$ matrix of predictive variables $\mathbf{X}$ in a beta regression model, which is tailored to model variables in the interval $(0, 1)$.¹⁴ Following Ferrari and Cribari-Neto (2004), the beta regression model assumes that the dependent variable is beta-distributed, with density

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\phi\mu) \Gamma(\phi(1-\mu))} y^{\phi\mu-1} (1-y)^{\phi(1-\mu)-1} \quad \text{for } 0 < y < 1 \quad (2)$$

with $E[y] = \mu$ and $Var[y] = \frac{\mu(1-\mu)}{1+\phi}$ and $\Gamma(\cdot)$ the gamma function. Using this notation, the standard beta regression model of Ferrari and Cribari-Neto (2004) also assumes independent realizations of $y_i \sim \mathcal{B}(\mu_i, \phi_i)$, and a linear regression model for the transform of the mean parameter. In this paper, we consider a *logit* link function for μ_i and a constant ϕ_i . This results in

$$\mu_i = \frac{e^{x_i^T \boldsymbol{\beta}}}{1 + e^{x_i^T \boldsymbol{\beta}}} \quad \text{and} \quad \phi_i = \phi. \quad (3)$$

Given that μ_i and ϕ_i are functions of the coefficients vector $\boldsymbol{\beta}$ and the scalar ϕ , respectively, we can estimate these parameters by maximum likelihood using a nonlinear optimization method. As a robustness check, we also consider the linear regression model,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}. \quad (4)$$

This model can be easily estimated using OLS. However, RRs are defined in the closed interval $[0, 1]$, and therefore predictions arising from a linear regression framework may lead to recovery rates that are negative or greater than unity.

A.2 Gradient boosted regression trees and random forests

Tree-based methods can identify clusters of data with similar properties, to capture nonlinear relationships between the target and its predictive variables as well as to reproduce step-like relationships. Ensembles combine weak learners¹⁵ to decrease the variance of a single prediction, therefore improving forecasting performance. *Bagging* and *boosting* are two popular types of ensemble methods.

Bagging (Breiman, 1996) forms multiple versions of a model by making bootstrap replicates of the learning set and using these as new learning sets. Then, the aggregate prediction is formed by averaging the forecast over the versions. Random forests are an extension of bagged regression trees¹⁶ However, bagged models may suffer from the risk of facing high correlations among individual models in the ensemble. Random forests (Breiman, 2001) is an ensemble of regression trees that additionally to standard *bagging*, also mitigates the risk of facing high correlations among individual trees in the ensemble by

¹⁴Following Smithson and Verkuilen (2006), we consider the transformation $\frac{y(n-1)+0.5}{n}$ where n is the sample size to adapt the beta regression framework to dependent variables assuming also the extremes 0 and 1.

¹⁵A weak learner is any machine learning algorithm that provides an accuracy slightly better than random guessing.

¹⁶Bagged regression trees are an ensemble learning method where multiple versions of regression trees are aggregated via bagging.

randomly selecting only a subset of predictors at each split of random forests. In this work, we consider an ensemble of 5,000 trees¹⁷ and tune the number of randomly selected predictors via 5-fold cross-validation.

Boosting also uses multiple versions of the same models but differs from bagging in that the weak learners are trained independently. In boosting, models are trained sequentially and adaptively. *Adaptively* means that the second version of the model tries to correct the errors present in the first iteration. However, to avoid over-fitting, only a percentage of each fitted value (called *learning rate*) is subtracted from the residual from the previous learner. This procedure is continued and models are added until a stopping condition is met. The number of boosting iterations, i.e., the number of trees, the learning rate, and the individual trees' depth are the main hyper-parameters for this model. We employ the stochastic gradient boosting of Friedman (2002) that improves both the approximation accuracy and execution speed while mitigating the risk of overfitting. This is done by randomly drawing (without replacement) a sub-sample of the training data from the full training dataset at each iteration and then using this one in place of the full sample to fit the base learner and compute the model update for the current iteration.

We also consider the XGBoost implementation of gradient-boosted trees (Chen and Guestrin, 2016). Similarly, to the stochastic gradient boosting of Friedman (2002), the number of trees, the learning (or shrinkage) rate, and the individual trees' depth are the main hyper-parameters. However, to further penalize complex model structures, mitigating the risk of overfitting, XGBoost uses L1 and L2 regularization on leaf weights.

A.3 Multivariate adaptive splines

The multivariate adaptive regression splines algorithm (MARS) (Friedman, 1991) is a non-parametric regression model that considers piece-wise transformations of the original predictive variables. Specifically, given a set t of cut points and p predictors, each predictor x_j is transformed into a set of *reflected pairs* obtained by:

$$(x_j - t_j)_+ = \begin{cases} x_j - t_j, & \text{if } x_j > t_j \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad (t_j - x_j)_+ = \begin{cases} t_j - x_j, & \text{if } x_j > t_j \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where $j = (1, 2, \dots, p)$ and $t_j \in \{x_{1j}, x_{2j}, \dots, x_{nj}\}$. To determine the cut points, a linear regression model is estimated following a greedy procedure on the reflected pairs for each predictor. The reflected pairs that achieve the smallest mean square error are then used for the model. The number of selected pairs for each predictor defines the *degree* of the MARS algorithm. A backward *pruning* procedure handles over-fitting by dropping the features that are associated with the smallest error rate when excluded. In this study, we consider MARS with a degree 1, 2, and 3, while the number of terms to be retained in the final model can range between 5 and 100. The degree and pruning parameters are estimated via 5-fold cross-validation. We also consider a bagged version of the MARS algorithm. The ensemble is made of 30 versions of the MARS algorithm trained with a degree of 1 or 2, while the number of terms to be retained in the final model can range between 5 and 100. These parameters are tuned using 5-fold cross-validation.

¹⁷Following Probst and Boulesteix (2017), we set this number to a computationally feasible large number, while still verifying that its cross-validated MSE does not exceed the ones obtained by smaller forests.

B Pseudo R code for hyperparameters tuning and forecast combination methods

In this section, we provide the details of the estimation of the methods under analysis. Let `mydf_training` and `mydf_fc` be the training set and the data set on which the forecasts combination are estimated, respectively. Let y be the target vector, X the matrix of predictive variables, F be the matrix of available forecasts, and FE the corresponding forecast errors. The following pseudo-code replicates the *COSE*⁺ optimal and robust combination scheme introduced by Roccazzella et al. (2022) and the nonlinear combination schemes employed in this study. All models were trained befitted from parallel computing to improve computational time and final predictions have been capped in the $[0, 1]$ interval. Due to computational limitations, we fixed hyperparameters from the first iteration.

B.1 Training of RF, GBRT, MARS and B-MARS models

```
##### Stochastic Gradient Boosting #####

gbm_grid <- expand.grid(.interaction.depth = seq(5, 80, by = 5),
.n.trees = c(10, 50, 100, 200),
.shrinkage = c(0.001, 0.01, 0.1, 0.2, 0.3, 0.5)
.n.minobsinnode=10)

gbm_fit <- train(as.formula("y~."),
data = mydf_training,
method = "gbm",
tuneGrid = gbm_grid,
distribution = 'gaussian',
trControl = trainControl(method = "cv",
number = 5))

##### XGB Trees #####

ALPHA <- seq(0, 1, by = 0.05) L1 regularization
LAMBDA <- seq(0, 1, by = 0.05) L2 regularization
ETA <- seq(0, 0.5, by = 0.05) Learning Rate
GAMMA <- seq(0, 0.5, by = 0.05) Minimum Loss for a New Leaf
SUBsamp <- seq(5, 1, by = 0.1) Row Sampling

# Iterate over the grids of hyperparameters for ALPHA, LAMBDA, ETA, GAMMA
# and SUBsamp and take the specification that minimizes the RMSE

xgbTrees.cv <- xgb.cv(seed = 42,
max_depth = 200,
```

```

eta = ETA[i],
lambda = LAMBDA[j],
alpha = ALPHA[z],
gamma = GAMMA[t],
subsamp = SUBsamp[h],
nthread = 6,
nrounds = 50,
objective = "reg:squarederror",
data = xgb_train,
nfold = 10,
showsd = TRUE,
stratified = TRUE,
metrics = "rmse",
verbose = TRUE,
print_every_n = 1L,
early_stopping_rounds = TRUE,
maximize = FALSE )

# Best tune for individual loans
xgbTrees <- xgboost(data = xgb_train, objective = "reg:squarederror",
max_depth = 200,
eta = 0.1,
lambda = 1,
alpha = 0.62,
gamma = 0.01,
min_child_weight = 10,
subsample = 0.9,
nthread = 6, nrounds = xgbTrees.cv$best_iteration)

##### Random Forest #####
rf_grid <- expand.grid(.splitrule= 'variance',
.mtry = seq(5, 40, by = 5),
.min.node.size = c(10))

rf_fit <- train(as.formula("y~."),
data = mydf_training,
method = 'ranger',
num.trees = 5000,
tuneGrid= rf_grid,
importance = 'impurity',
trControl = trainControl(method = "cv",
number = 5))

```

```
##### MARS #####
mars_grid <- expand.grid(degree = 1:3,
nprune = 1:100)

MARS_fit <- train(as.formula("y~."),
data=mydf_training,
method = "earth",
metric = "RMSE",
tuneGrid = mars_grid,
trControl = trainControl(method = "Cv",
number = 5))

##### Bagged MARS #####

BMARS_fit <- train(as.formula("y~."),
data=mydf_training,
method = "bagEarth",
B = 30,
metric = "RMSE",
tuneGrid = mars_grid,
trControl = trainControl(method = "Cv",
number = 5))
```

B.2 Minimum variance combination of forecasts with non negative weights

Using the R package `quadprog`¹⁸, the function `Weights_P` estimates the combining weights given the covariance matrix `Qmat`.

```
Weights_P <- function(Qmat){
m <- ncol(Qmat)
Eq <- ones <- rep(1,m)
Ineq <- diag(1, m)
b <- c(1, rep(0,m))
dvec = as.matrix(rep(0, m), nrow = m)
Amat = t(rbind(Eq,Ineq))
bvec = as.matrix(b, nrow =m)
meq = 1
Weights <- solve.QP(Qmat, dvec, Amat, bvec, meq)
return(Weights$solution)}
```

Opt + Constrained Optimization.

¹⁸<https://cran.r-project.org/package=quadprog>.

```
S <- cov(FE)
Weights_S <- Weights_P(S)
```

B.3 Non linear combination schemes

Using the R library `caret` (Kuhn, 2008) and its dependencies, we can estimate NN_{FC} . The combined forecast can be simply obtained with `predict`.

```
nn_grid <- expand.grid(decay = seq(0, 0.3, by = 0.01),
size = c(10, 20, 30, 50, 100), bag = TRUE)

nn_fc <- train(as.formula("y~."),
data=mydf_fc,
method = "avNNet",
tuneGrid = nn_grid,
maxit = 5000,
MaxNWts = 800,
repeats = 20,
trControl = trainControl(method = "Cv",
number = 5))
```

C Distribution of recovery rates on pools

Let N_i and R_i be the outstanding amount (i.e., amount to recover) and recovery rate associated with credit i , respectively. The average recovery rate on a pool of defaulted credits $1, \dots, n$ is

$$\langle R_n \rangle = \frac{\sum_{i=1}^n N_i R_i}{\sum_{i=1}^n N_i} = \sum_{i=1}^n w_i R_i, \quad w_i = N_i / \sum_{j=1}^n N_j.$$

We consider a one-factor latent variable model, i.e., the R_i 's are modeled as random variables taking values in $[0, 1]$ and being mutually independent conditional upon a single systematic risk factor, Z . It is well known from the strong law of large numbers that, conditional upon Z an under technical assumptions¹⁹, the variable $\langle R_n \rangle$ converges to its conditional expectation $\langle R_n(Z) \rangle := \mathbb{E}[\langle R_n \rangle | Z]$ as $n \rightarrow \infty$ (see Gordy, 2003, for details). This result essentially says that as the exposure share w_i of each asset in the portfolio goes to zero, idiosyncratic risk in portfolio loss is perfectly diversified away, so that the average recovery rate becomes a deterministic function of the systematic factor, Z . We conclude from this result that the distribution of $\langle R_n \rangle$ can be approximated reasonably well by that of $\langle R_n(Z) \rangle$, for n large enough.

For the sake of illustration, consider an equally-weighted homogeneous pool, with $w_i = 1/n$ and where the recovery rates R_i are modeled as Bernoulli random variables with parameter $\mu = \mathbb{P}(R_i = 1)$. In this illustration, we consider the one-factor Gaussian copula dependence scheme, which is the corner stone

¹⁹including that $\sum_{i=1}^n N_i \rightarrow \infty$ and $\max_{i \in \{1, 2, \dots, n\}} w_i = O(n^{-(1/2+\zeta)})$ for some $\zeta > 0$ as $n \rightarrow \infty$.

of the loss model adopted in the Basel regulation on capital requirements.²⁰ Noting $1_{\{\cdot\}}$ the indicator function,

$$R_i = 1_{\{X_i \leq \Phi^{-1}(\mu)\}} \quad \text{where} \quad X_i = \sqrt{\rho}Z + \sqrt{1-\rho}\tilde{X}_i$$

and $Z, \tilde{X}_1, \dots, \tilde{X}_n$ are i.i.d. standard Normal variables with cumulative distribution function Φ . Then, from the linearity of the conditional expectation operator,

$$\langle R_n(Z) \rangle = \sum_{i=1}^n w_i \mathbb{E} [1_{\{X_i \leq \Phi^{-1}(\mu)\}} | Z] = \frac{1}{n} \sum_{i=1}^n \mathbb{P}(X_i \leq \Phi^{-1}(\mu) | Z) = \Phi \left(\frac{\Phi^{-1}(\mu) - \sqrt{\rho}Z}{\sqrt{1-\rho}} \right).$$

This leads to the following approximation for the distribution of $\langle R_n \rangle$ for large n :

$$\mathbb{P}(\langle R_n \rangle \leq x) \approx \mathbb{P}(\langle R_n(Z) \rangle \leq x)$$

where, using $\Phi(x) = 1 - \Phi(-x)$,

$$\begin{aligned} \mathbb{P}(\langle R_n(Z) \rangle \leq x) &= \mathbb{P} \left(\Phi \left(\frac{\Phi^{-1}(\mu) - \sqrt{\rho}Z}{\sqrt{1-\rho}} \right) \leq x \right) \\ &= \mathbb{P} \left(Z \geq \frac{\Phi^{-1}(\mu) - \sqrt{1-\rho}\Phi^{-1}(x)}{\sqrt{\rho}} \right) \\ &= \Phi \left(\frac{\sqrt{1-\rho}\Phi^{-1}(x) - \Phi^{-1}(\mu)}{\sqrt{\rho}} \right). \end{aligned}$$

Noting φ the density of the standard normal random variable, the asymptotic approximation for the averaged recovery rate density is

$$\hat{f}_{\langle R_n \rangle}(x) := \frac{1}{\varphi(\Phi^{-1}(x))} \sqrt{\frac{1-\rho}{\rho}} \varphi \left(\frac{\sqrt{1-\rho}\Phi^{-1}(x) - \Phi^{-1}(\mu)}{\sqrt{\rho}} \right) \quad (6)$$

Figure 12 illustrates the true and asymptotic distributions of the averaged recovery rate on an equally-weighted homogeneous pool in a single factor model and Gaussian coupling scheme with $\mu = \mathbb{E}[R_i] = 0.2$ for various values of ρ and n . The exact distribution of $\langle R_n \rangle$ is discrete, and the corresponding probability mass function is obtained by noting that conditional upon Z , the average recovery rate follows a binomial law. The unconditional law of $\langle R_n \rangle$ is found by integrating Z out, which is performed numerically, using Gaussian quadrature. The asymptotic density in eq. (6) offers a good approximation of the exact distribution associated with the recovery rate model (in practice, as from n is equal to a few tenths).

²⁰In Basel regulation, the goal is to model the default events, i.e., the default indicators. However, in this setup where the recovery rates are assumed to be binary, a similar approach can be used to couple the recovery rates. Indeed, the recovery rates are known to be negatively correlated with the default rate, hence, exhibit a positive dependence.

Figure 12: Exact (blue dots) and asymptotic (red solid line) distributions of the average recovery rate on an equally-weighted homogeneous pool ($w_i = 1/n$, $\mu = 0.2$) of size n in a single factor Gaussian model for $n = 10, 50, 1000$ (left to right). The asymptotic distribution provides a good approximation even for relatively small n (a few tenths). Note that the density curve (in red) is rescaled so that its maximum matches with the maximum of the probability mass function of $\langle R_n \rangle$.

