Business cycle and realized losses in the consumer credit industry

Walter Distaso[1] Francesco Roccazzella[2,*] Frédéric Vrins[3,4]

[1]Imperial College Business School, United Kingdom
[2]IESEG School of Management, France
[3]UCLouvain, LIDAM - LFIN, Belgium
[4]HEC Montréal, Department of Decision Science, Canada

July 25, 2023

Abstract

We analyze the determinants of loss given default (LGD) of consumer credit. Previous studies conclude that only contract-specific variables such as exposure-at-default or age matter; in contrast with defaulted corporate bonds, macro variables like GDP or business cycle are irrelevant. Exploiting a much larger dataset including more than 6 million of Italian consumer debt from 2007 to 2019, considering more than 40 predictors and a variety of models covering standard regressions, machine learning techniques, and forecast combination schemes, we consistently obtain opposite results. We find that regional social and welfare indicators and credit cycle variables contribute to enhancing forecasting performance by up to 15 percentage points in terms of R^2. The relationships between the expected LGD and the macro predictors unveiled by accumulated local effects plots confirm 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, too. In addition, our analysis suggests that the most effective approach for predicting LGDs for both individual credits and portfolio of credits is to use forecast combination schemes. Among these methods, forecast combinations based on artificial neural networks stand out. Using model confidence set theory, we find that they consistently outperform the other considered models in terms of mean absolute error and root mean squared error, always joining the superior set of models. Our results are important for credit institutions, third-party debt collection firms, or regulatory bodies in charge of capital requirements policies.

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

[*]Corresponding author - *Address:* Voie du Roman Pays 34, B-1348 Louvain-la-Neuve, Belgium. Email: francesco.roccazzella@uclouvain.be

1 Introduction

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

Consumer credit is a fast-growing industry worldwide. In September 2022, for instance, consumer debt rose to USD 4,700.821 bln in the US and to EUR 715.106 bln in the euro area. This represents a growth of 76.6% in the US and 11.74% in the euro area concerning levels observed before the outburst of the great financial crisis in September 2008. In the euro area, it accounts for 10.8% of the outstanding credit for households.[1]

Consumer credit is primarily concerned with credit risk, which arises when consumers fail to repay their debt, resulting in losses for creditors. The degree of loss suffered by the creditors can be measured using two parameters, namely the exposure at default (EAD) and the recovery rate (RR) (or 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, whereas the RR determines the magnitude of the losses incurred by creditors when closing the debt collection process.

Confronted with a non-performing credit (NPC), a debt holder has four 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, to 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 RRs are likely to be.

Although a vast literature deals with the RRs of defaulted corporate bonds, little is known about the RRs of consumer credits. However, it is important to note that these two types of recovery rates are not easily comparable. For instance, the assets are very different: bonds are issued by large institutions, hence, are much more liquid than individual or SME loans. In addition, both the amount and the nature (public vs private) of available information about them are different. Finally, it is well-known that the empirical distribution of the RRs of defaulted corporate bonds is smooth and unimodal, while that of NPCs exhibits a pronounced bimodal pattern (see, e.g., Gambetti et al., 2019; Nazemi et al., 2022). These differences suggest that one can hardly extrapolate the findings related to defaulted corporate bonds to NPCs. As detailed in the next section, most of the studies on defaulted

[1] Data 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/.

corporate bonds and consumer credit RRs highlight that contract-specific variables (such as the seniority or the sector for bonds, and the maturity or the principal for consumer credit) matter. However, 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 is known since the seminal paper of Altman et al. (2005) that realized recovery rates of corporate bonds are negatively correlated with the default rate, showing that macroeconomic variables matter for these assets. Surprisingly enough, this would not apply to consumer credit, where macroeconomic variables were shown to be empirically irrelevant. This contrasts with the idea that income and the residual value of consumers' personal assets depend on the environment in which the recovery process takes place. In such a case, if realized recovery rates display pro-cyclical dynamics, consumer debt owners will risk facing greater realized losses than expected.

In our view, four main reasons can be put forward to explain this paradox. First, it is very difficult to access a meaningful dataset: data is often scarce and limited to a specific sector of the consumer credit industry, preventing the findings to be easily generalized. Second, the samples under analysis are relatively short in the time dimension while the economic indicators are low-frequency time series that evolve across business cycles. The low variability of MS indicators in short samples may explain their marginal relevance for describing consumer debt RRs compared to more granular credit-specific features such as exposure at default or debtor's age. Third, researchers have 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. Fourth, recovery rate modeling has mostly focused on in-sample *vs* out-of-sample splits to analyze recovery rate determinants or discuss the best predictive strategies. Nevertheless, when the data is randomly assigned to in-sample and out-of-sample sets, realizations of RRs used for training the models have likely occurred *after* the defaults of credits used for evaluating the models. The underlying assumption is that the data-generating process is time-invariant which, if not met, can potentially lead to a look-ahead bias in the predictions, therefore invalidating the findings.

This paper investigates the relevance of macroeconomic and social variables in explaining recovery rates and therefore realized losses in the broader consumer credit industry by relying on a real portfolio of NPCs owned by a debt collector.

We make the following contribution to the literature on consumer credit. First, we analyze a database counting more than 6 million NPCs from utilities, banking, financial and commercial sectors in Italy, spanning the years 2007 - 2019. To the best of our knowledge, this is the largest dataset considered so far in the academic literature. Second, we study the determinants of LGDs both at the individual NPC levels but also at portfolio levels. While existing studies focus on the former, the latter also matters when, for example, the problem at hand is to determine the fair price of a portfolio of NPCs to be transferred from one party to another. Third, exploiting the heterogeneous economic dynamics across regions and provinces of Italy[2], the data permits us to understand how the economy impacts LGDs in the consumer credit industry. In contrast with previous studies analyzing the determinants of NPC using smaller datasets focusing on specific types of NPC, we find that excluding MS indicators from the set of predictors significantly harms forecasting performance both at individual credit and portfolio levels. We believe that this effect comes from the lack of variability in macro indicators when considering smaller time frames. 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, when it comes to time-consistent prediction exercises, our conclusion becomes even more significant. Discarding the time dimension, including MS variables improves the R^2 for beta and linear regression models by 4 percentage points. In the inter-temporal framework, this number jumps to 15 percentage points. Accumulated local effect plots (ALE) confirm that our results are in line with 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, while Section 4 outlines the forecasting methods used in our study. We also describe our methodology for data sampling in Section 5. Section 6 presents our results, and we conclude the paper in the final section.

2 Literature review

Extensive research has been conducted on the factors influencing RRs. In this section, we begin by examining theoretical models that explain the factors that determine RRs, while also highlighting potential distinctions between corporate bonds and consumer credit. Subsequently, we investigate empirical evidence that quantifies the significance of macroeconomic predictors in predicting recovery rates, comparing the findings in the context of defaulted corporate bonds and consumer credit.

[2]For instance, in 2018, the GDP per capita ranged from EUR 23,879 in Sicily to EUR 35,968 in Lombardy, with the average for Italy being EUR 31,641. Similarly, during the first half of 2020, credit to firms increased by varying percentages in different regions, with a higher increase in the Center, North-West, and North-East regions and a lower increase in the Islands and South of Italy, according to (Bank of Italy, 2020).

The recovery rate for corporate loans and bonds depends on the value stream that can be obtained from liquidating a firm's assets in the event of default. The expected residual value to claimants is greater when the firm's value is higher and its level of indebtedness is lower (Merton, 1974). The macroeconomic and financial conditions that firms and their peers operate in can affect the value stream generated by the firm's assets, as well as the revenues and costs associated with bankruptcy proceedings in the case of default, thus directly impacting the RR (Shleifer and Vishny, 1992). This idea has been supported by extensive empirical evidence, with studies such as Altman et al. (2005); Acharya et al. (2007); Bruche and González-Aguado (2010); Gambetti et al. (2019) highlighting the importance of macro-financial conditions in explaining the determinants of the recovery rate of corporate bonds.

Conversely, little is known about consumer credit recovery rates, including whether they display pro-cyclical patterns similar to those observed in corporate bonds or consumer default rates (Nakajima and Ríos-Rull, 2014; Livshits, 2015). However, from a theoretical perspective, we can expect that the MS conditions should influence similarly the recovery rate for consumer credit. The income and residual value of consumers' personal assets are contingent on the environment in which the recovery process takes place. However, differences in access to information and legal frameworks make recovery procedures for consumer debt and corporations inherently different. While firms are legally obligated to periodically disclose their balance sheets and income statements publicly, accessing such information for final consumers and small and medium enterprises (SMEs) can be costly or even impossible. This poses a significant obstacle when pursuing credit recovery, especially in the case of unsecured consumer credit. The legal environment may also restrict access to information and limit actions to pursue recovery. For example, in the credit card industry, Fedaseyeu (2020) found that consumer protection legislation governing third-party debt collection reduces the number of third-party debt collectors and increases the loss given default on delinquent credit card loans.

In contrast to corporate bonds, the empirical evidence on the importance of MS indicators for explaining consumer credit RR is inconclusive. Thus far, the results appear to vary based on the time in which the studies were conducted, the specific type and sector of the NPC, and the number of MS variables that were considered.

For example, the study by Bellotti and Crook (2012) focused on defaulted credit-card accounts from 1999 to 2005 and found that including macroeconomic variables may enhance the predictive power of recovery rates for defaulted credit-card accounts. They consider three macroeconomic series and found that higher interest rates and unemployment levels at the time of default are associated with lower recovery rates, while higher earnings growth leads to better recoveries. Conversely, Leow et al. (2014) find that macroeconomic variables may improve the predictive performance of mortgage loan RR estimates, but not of personal loan RR. They consider a broader set of economic variables at the UK national level, 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 and unsecured personal banking loans ranging from 1990 to 2002 and from 1989 to 1999, respectively.

Looking at the broader market of consumer credit, Thomas et al. (2012) compared debt characteristics and the RR of consumer credit for in-house and third-party collection policies. However, their analysis does not consider macroeconomic and financial predictors and it only considers defaulted consumer credits over two years in the 1990s (for in-house collection) and 2000s (for third-party collection). Beck et al. (2017) examined the RR of consumer credits for goods and services in the period 2004–2008 in Germany, but mainly focused on idiosyncratic determinants such as the EAD or prior debtor-specific collection rates. Only the growth rate of gross domestic product and the unemployment rate in the debtor's federal state for the year in which the account was handed over to a third-party collection agency were considered. They have mixed results on the relationship between macroeconomic conditions and RR, with only the unemployment rate having a consistently negative effect on the recovery rate. Nazemi et al. (2022) consider 65,535 defaulted unsecured consumer credits purchased between 2010 and 2013 from a German telecommunications company. To study the effects on the recovery rate related to the economic conditions, they include the unemployment rate at the province level and *excessive indebtedness rate*, which measures the percentage of adults with strong negative credit ratings in a specific province. However, Nazemi et al. (2022) find that including such variables does not improve the prediction accuracy.

Finally, as underlined by Kalotay and Altman (2016) and Nazemi et al. (2022), evidence in the relevance of MS and NPC-specific predictors can be faulted. Recovery rate modeling has mostly focused on in-sample *vs* out-of-sample splits to analyze recovery rates determinants (Gambetti et al., 2019) or discuss the best predictive strategies (Nazemi and Fabozzi, 2018; Nazemi et al., 2018, 2022; Gambetti et al., 2022). When the data is randomly assigned to in-sample and out-of-sample sets, realizations of RRs used for training the models have likely occurred after the defaults of credits used for evaluating the models. The underlying assumption is that the data-generating process is time-invariant which, if not met, can potentially lead to a look-ahead bias in the predictions, therefore invalidating the findings.

Contrasting to previous studies, with an unprecedented database at hand, these limitations can be overcome. We consider a broader range of consumer debt, including banking loans, financing contracts, sales credits, and debts related to telecommunication and utilities. We also consider a larger set of 13 MS variables to describe macroeconomic and social conditions. Moreover, by conducting an inter-temporal prediction exercise, which ensures that only sample points observed before the default event are used during the training process, we may dispel the doubt on whether the relevance of MS predictors found in a simple out-of-sample exercise is confirmed.

3 Data

To properly understand the data associated with the consumer credit business, it is important to highlight its specificity, underlining its main features compared to other types of credits. We then describe the data provided by the third-party collector agency and the set of economic indicators we use in our analysis.

3.1 Non-performing consumer credit (NPC) database

The database under study (called NPC DB hereafter) is private and has never been exploited in any publication so far. It counts 6,493,794 NPCs from Italy and contains defaulted consumer credits that the third-party collector is entitled to recover for a maximum duration period of 365 days since receiving the mandate from the original lender.

We observe eight numerical and categorical contract-specific variables such as 1) the realized recovery rate, 2) the total to recover (TtR), 3) the Principal-to-TtR ratio, which measures the importance of interest payments and ancillary fees relative to the TtR, 4) the industry originating the credit, 5) the debtor's fiscal identification number[3], 6) the postal code of its main residence, 7) the date when the third-party agency was authorized to recover the defaulted credit, and 8) the maximum duration granted to third-party collectors to recover defaulted debt.

We can now 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. The database offers a representative picture of the situation in the country as all 20 regions and 101 out of 107 provinces are represented.

Table 1 reports the descriptive statistics of the recovery rates across types of debtors, and sectors and conditional on the 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 information is missing for the remaining credits. The defaulted credits originate from the utility (35.04%) and telecommunication sectors (48.03%), from financing (car loans, credit cards, leasings, 5.45%), banking (overdrafts and personal credits, 3.72%) and commercial credit (sales credit, 7.78%).

[3] To guarantee debtors' anonymity; the fiscal identification number has been partially obscured, permitting us to extract data on whether the debtor is a professional or a physical person, and for this latter, its age.

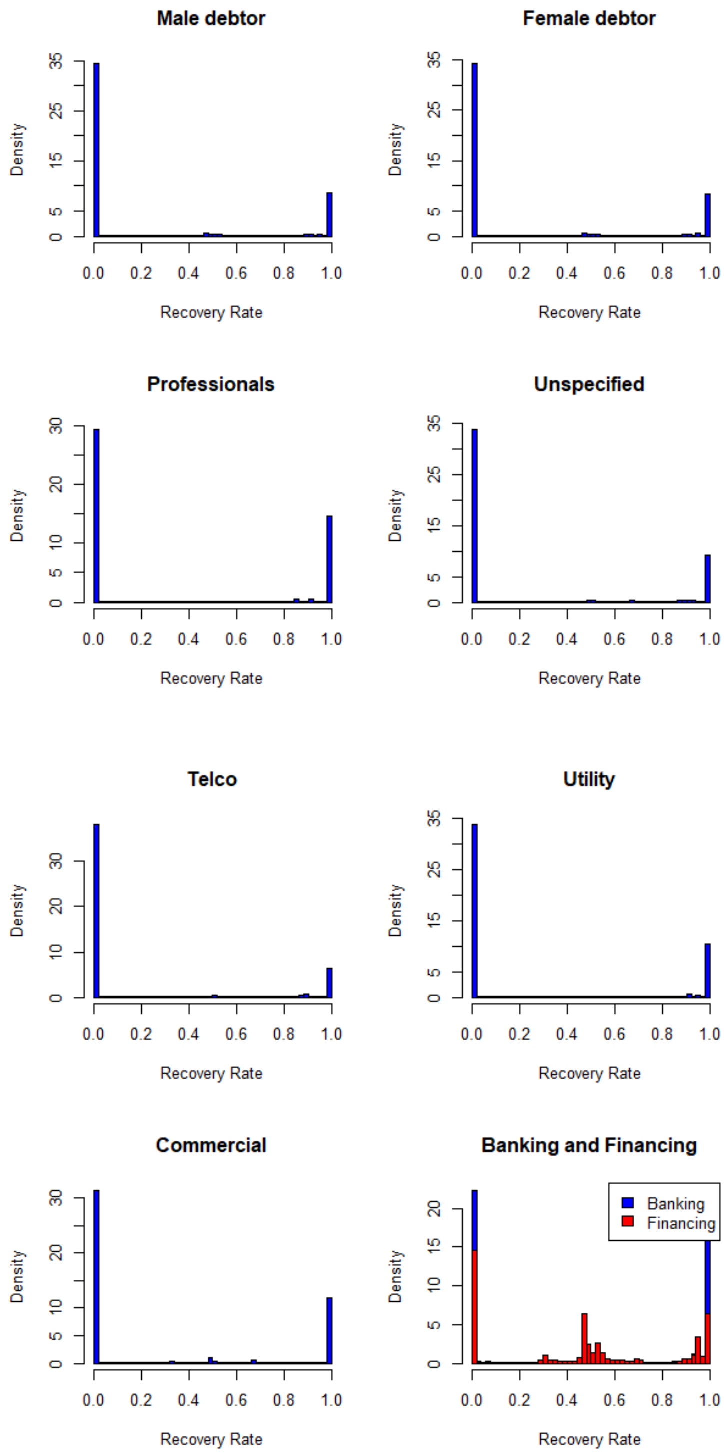
Table 1: Recovery rate descriptive statistics of non-performing credits by debtors type, sectors, and conditional on Total to Recover (TtR) ranging from EUR 50 to EUR 10,000.

				Percentile				
Debtor	# obs	μ	σ	.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	0	.396	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

We observe that 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% recovery rates, respectively. NPCs originating from the banking and financing industry display instead an average 48% and 47% recovery rate, respectively. 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 & finance sector, which exhibits a third mass around 50%. Moreover, Table 1 confirms previous evidence on the negative relationship between recovery rate and total amount to recover (Nazemi et al., 2022).

We count four categories of debtors, male and female physical debtors, professionals, and unspecified. In the case of the debtor being a physical person, we also observe the age. For professional and unspecified debtors, the variable *age* is set to its median in the

Figure 1: Distributions of the realized recovery rate of individual NPCs.



sample, i.e., 55 years old and the binary variable *professionals* is included. The data set includes information on the debtor's postal code and the date when the third-party agency was authorized to recover the defaulted credit. This information enables us to associate each debtor with a particular province and region at a specific point in time, which allows us to 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. We 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[4].

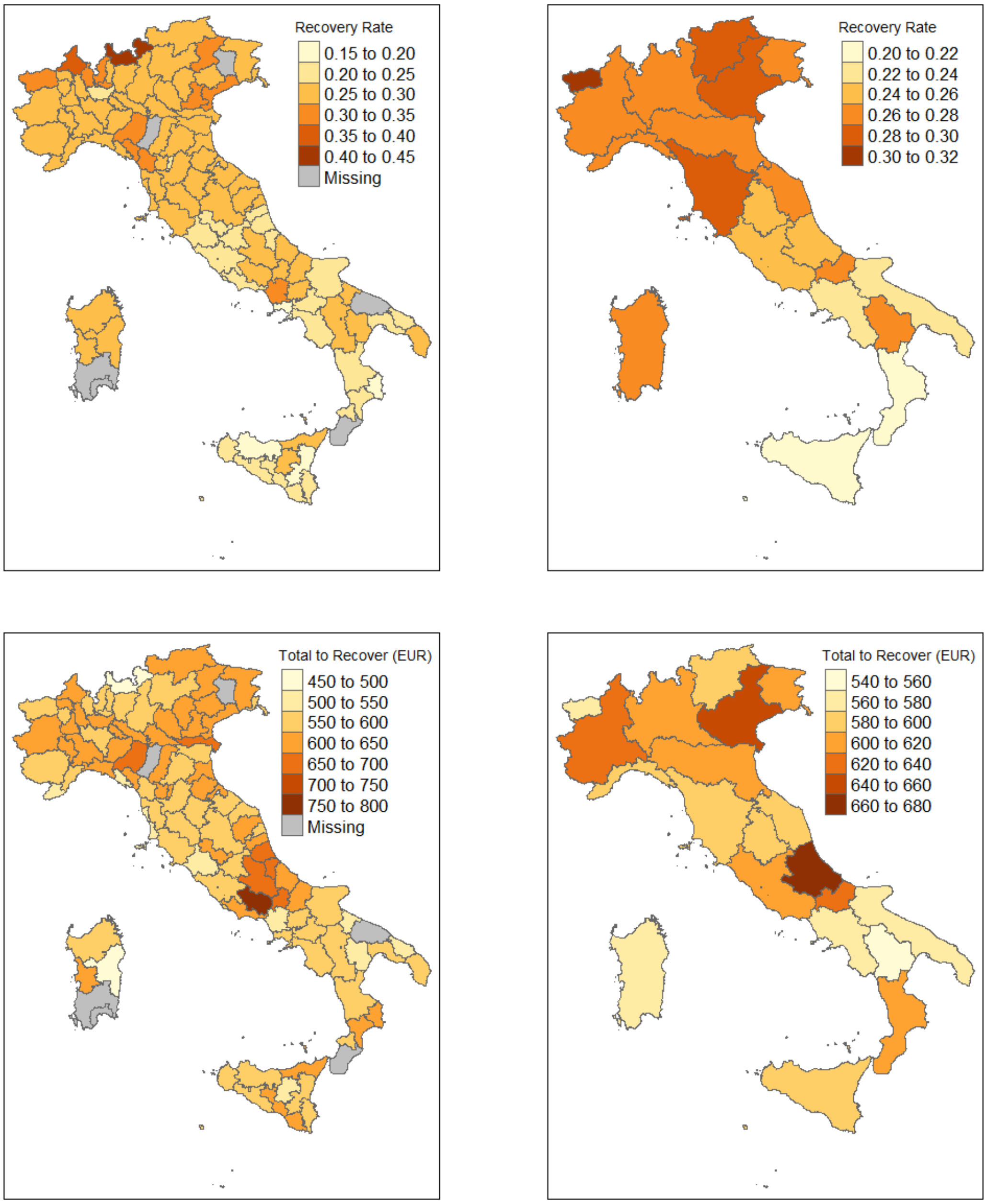
The left charts of Figure 2 display the average recovery rates and total recovery values for consumer NPCs in various Italian provinces, while those on the right-hand side show the same information but at the regional scale. Generally, the recovery rates for NPCs are on average higher in the northern 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. Contrasting to the trend in terms of recovery rates between the north and south, the bottom 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.

Figure 3 presents information on the average ratio of principal to the total amount to recover (top) and the average age of debtors (bottom). The data indicates that interest and ancillary fees have a relatively small impact on the total amount to be recovered, with the principal amount accounting for 86% to 96% of the total. This is particularly evident for NPCs originating from the center and south of Italy, where the principal accounts for over 90% of the total. Furthermore, 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 of 60-66 years.

In the top panel of Figure 4, we can 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.

[4]Information is available at: https://www.istat.it/en/archivio/227202

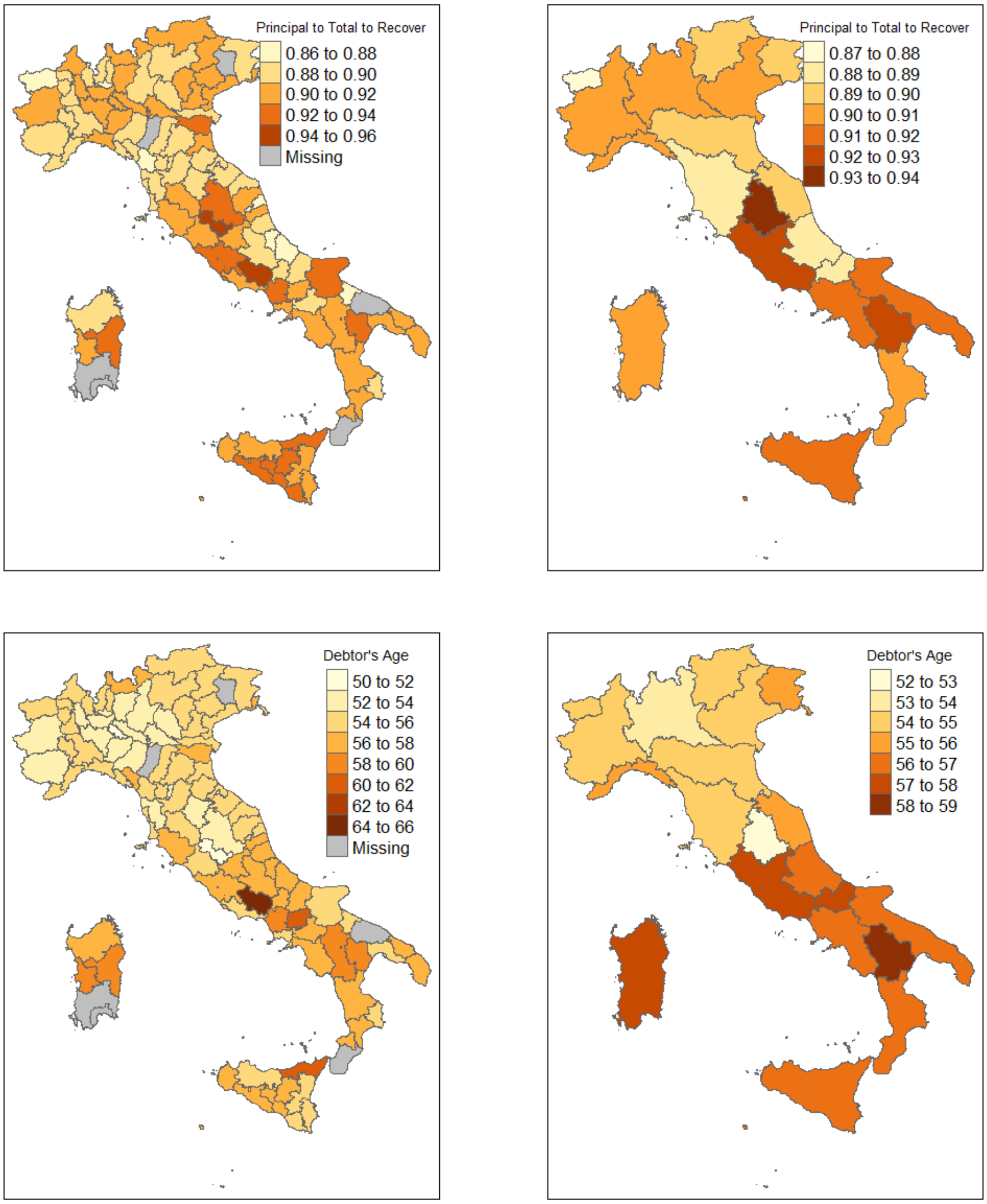
Figure 2: Heterogeneity of recovery rates (top) and total to recover amounts (bottom) at province (left) and region (right) level.



3.2 Economic 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 with macroeconomic and social indicators collected outside the NPC DB. We have a total of 13

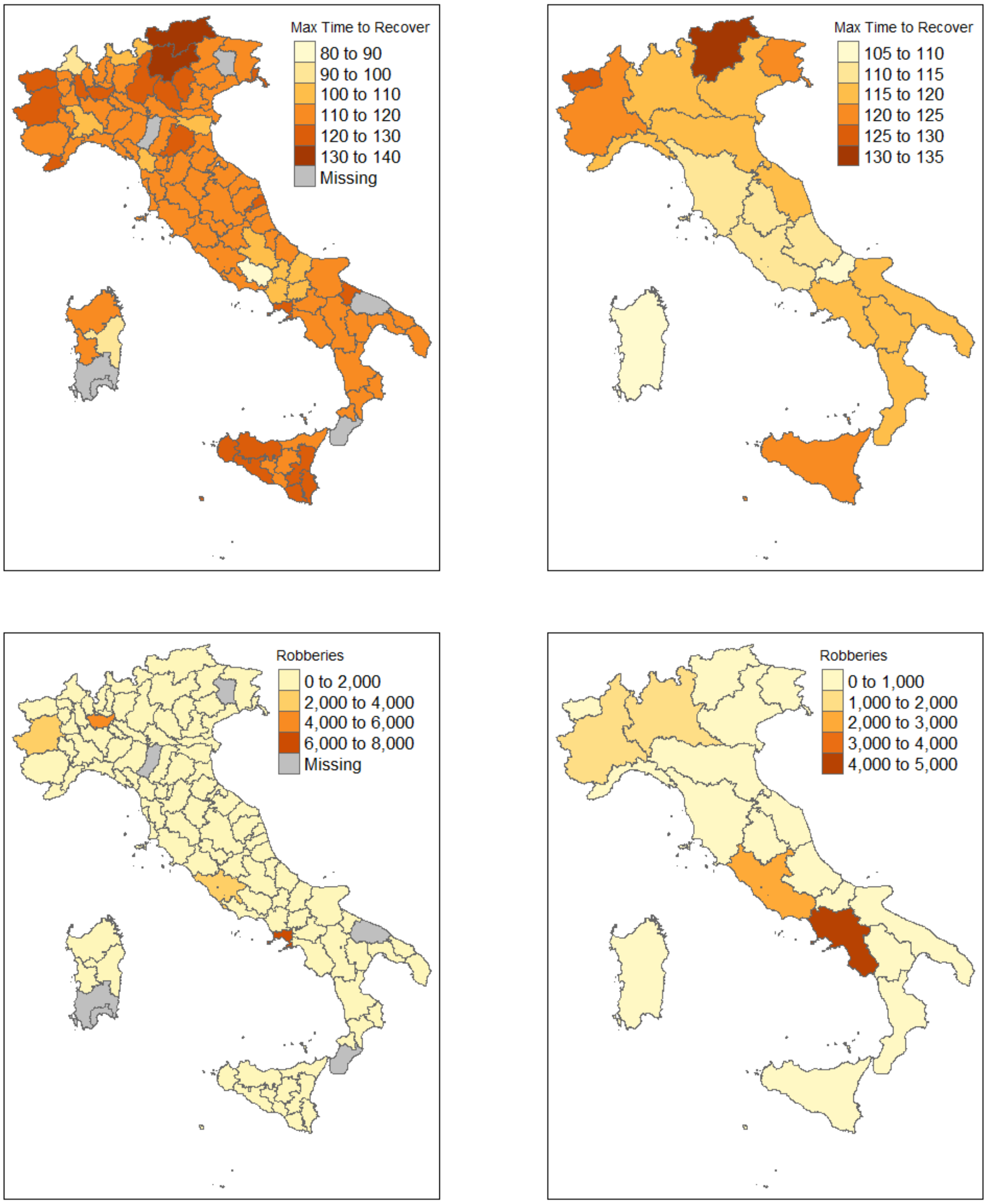
Figure 3: Heterogeneity of the principal to total to recover (top) and of the debtors' age (bottom) at province (left) and region (right) level.



variables that we can classify into four broad classes of macroeconomic and social indicators measuring the effects of the business cycle, credit cycle, and economic uncertainty but also the social environment on RRs.

We start with the four indicators of the business cycles. We consider the unemployment rate and the year-on-year growth rate of the Harmonized Index of Consumer Prices (HICP)

Figure 4: Heterogeneity of the maximum duration collection process (top) and of the number of robberies (bottom) at province (left) and region (right) level.

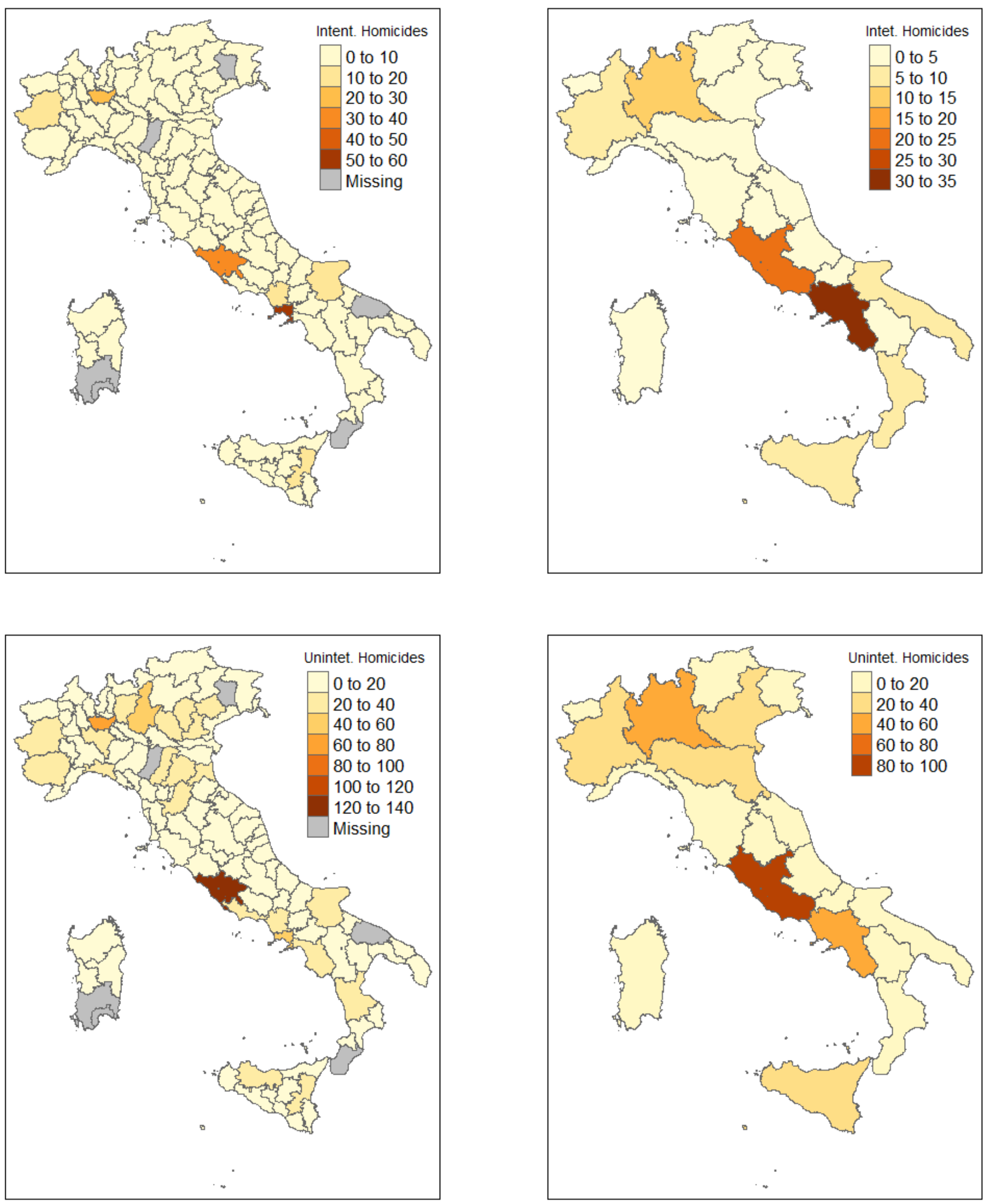


as a proxy for inflation.[5] Unfortunately, because of structural breaks in the reporting of inflation and unemployment series at the regional or provincial level, we stick with these measures at the national level. We overcome this limitation when dealing with real gross value added as a measure of actual economic activity. This data is obtainable at the

[5]Both series are monthly and seasonally adjusted.

regional level and on an annual basis. We also incorporate its yearly growth rate as a gauge of business cycle fluctuations. Figure 6 illustrates the average real gross value added and yearly growth rate, highlighting significant economic differences between regions in terms of growth and levels.

Figure 5: Heterogeneity of the number of intentional (top) and unintentional (bottom) homicides at province (left) and region (right) level.

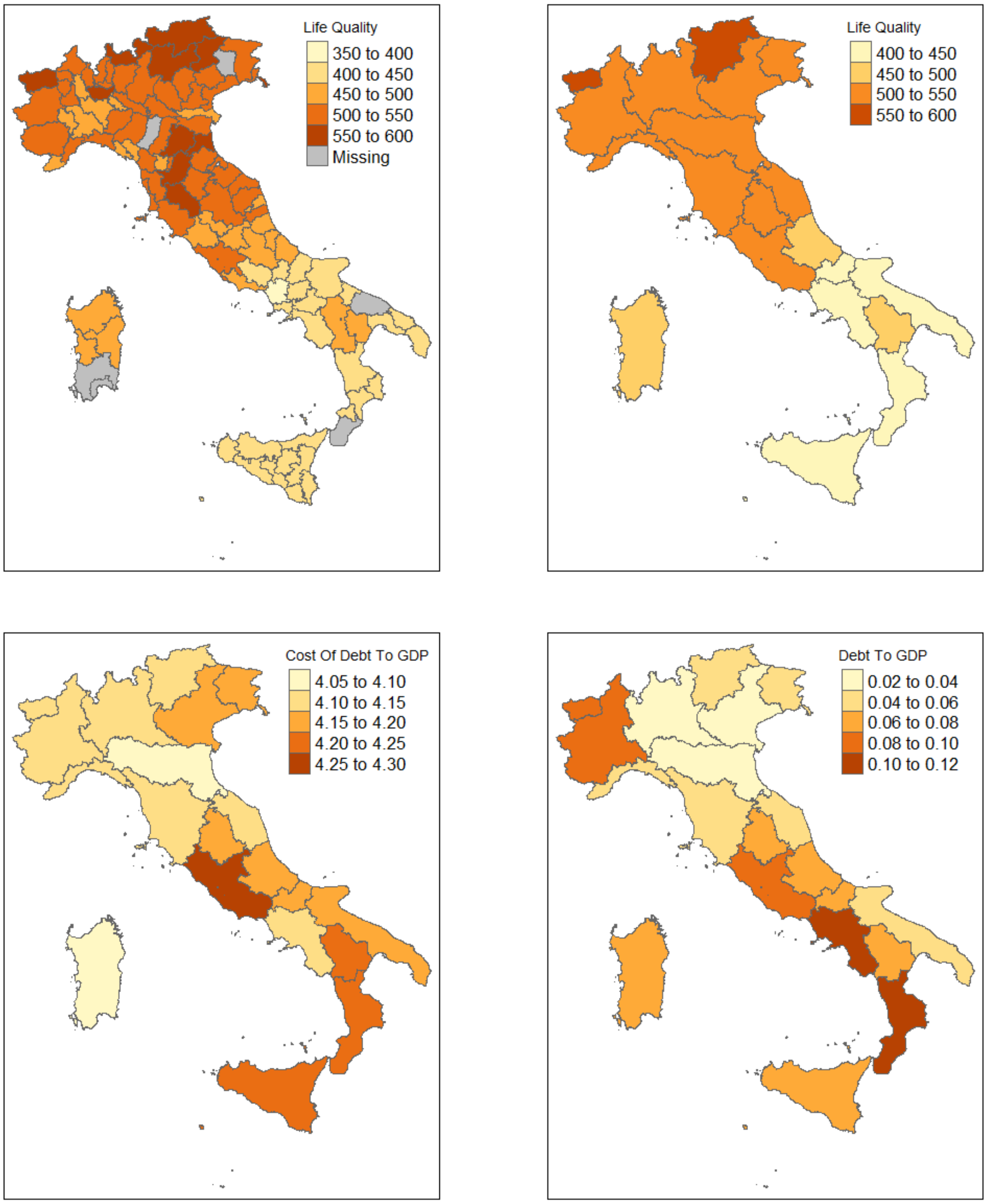


Then we move to the four variables used to proxy the dynamics of the credit cycle. The public debt and cost-of-debt to GDP ratios are used to evaluate credit conditions at both the national and regional levels. In Figure 5, the lower panel shows the variations in the cost-of-debt to GDP (on the left-hand 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 also incorporate the spread between Italian and German 10-year constant maturity sovereign bond yields and the cost of debt to GDP ratio for the Italian economy to control credit market conditions at the national level. The economic policy uncertainty index of Baker et al. (2016) for Italy is the last macroeconomic predictor we include.

The social environment can also significantly affect the outcome of the collection process. For instance, higher levels of criminal activity, lower income, or less developed areas may directly impact the collection process and affect the realized recovery rate. 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 per 100,000 inhabitants, and 4) the Life Quality Index published by the Italian financial newspaper Il Sole 24 Ore. This 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. The lower panel of Figure 5 and Figure 6 reveal that Milan, Rome, and Naples provinces have higher numbers of robberies, and intentional and unintentional homicides per 100,000 inhabitants compared to other provinces. The upper panel of Figure 7 displays 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).

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 Table 2 and 3.

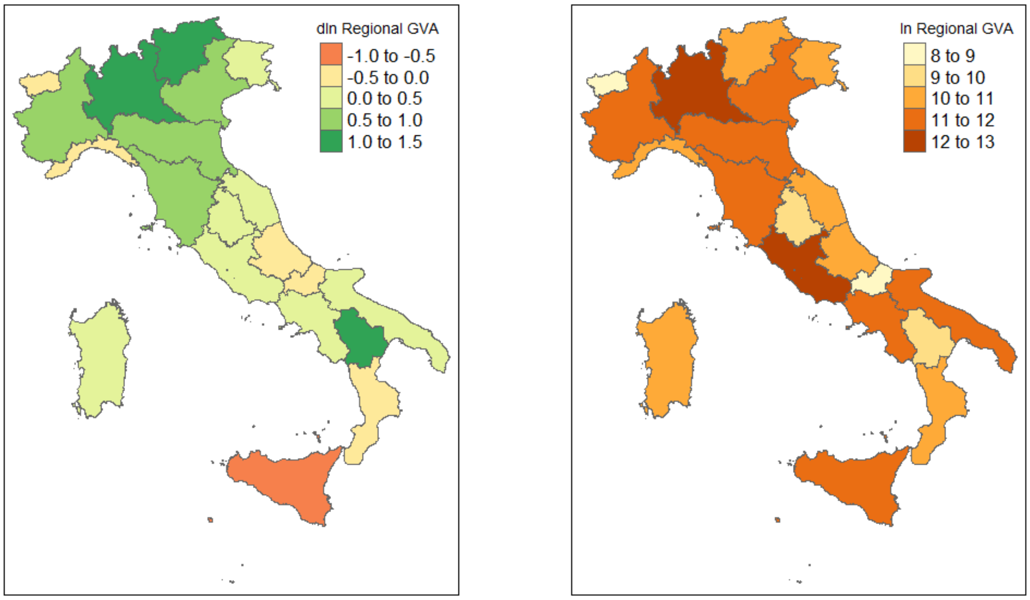
Figure 6: Heterogeneity of Life Quality, Cost of debt and Debt to GDP rations at province and region level.



3.3 Portfolios of defaulted consumer credits

Third-party collectors acquire *the right to recover* defaulted credit in lots by participating in auctions with information on the sector and the geographical area originating the pool of credits and the maximum duration granted to the collection process but very little information on the specific constituents of the lot. They handle the recovery process in the

Figure 7: Heterogeneity of GDP level (left) and growth (right) at region level.



name of the lender and earn a commission on the total amount recovered and on the number of collection actions performed. Therefore, studying the determinants at the portfolio level better reflects the third-party collectors' business model.

We now describe how we formed the portfolios of defaulted consumer credits we use for our analysis. For each quarter between (and including) 2007 Q1 and 2019 Q4, we randomly draw (without replacement) samples of 100 individual credits that have been acquired in that quarter. We consider portfolios of 100 credits because it 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. For each quarter, we continue to form portfolios till we have less than 100

Table 2: List of considered macroeconomic, financial, and social (MFS) variables.

Variable	Abrev.	Mean	Std.Dev.	Min	Pctl. 25	Pctl. 75	Max
Regional Gross VA Mln	R-GVA	119414.493	96610.849	4103.1	64286.2	141279	344629.2
Regional Gross VA dln1	R-GVA d1	0.108	1.975	-8.569	-0.258	1.149	8.947
National Unemployment	Un	11.807	5.872	2.747	7.339	18.382	23.415
EPU Italy	EPU	110.972	32.992	31.702	89.521	130.025	241.018
HICP YoY	HCIP	0.861	0.825	-0.509	0.193	1.363	4.165
Regional Debt to GDP	R-DtGDP	0.07	0.029	0.024	0.042	0.09	0.149
NationalDebt to GDP	N-DtGDP	129.787	9.979	103.9	132.5	134.8	135.4
National Cost of Debt to GDP	N-CtGDP	4.148	0.484	3.6	3.8	4.6	5.2
Spread ITA-GER 10Y bonds	Spread	171.549	65.719	21.066	130.701	205.274	460.318
Life Quality Index	LQ	490.15	66.603	359.97	425	545.55	641.49
Intentional Homicides	INHom	11.03	16.52	0	1	13	109
Unintentional Homicides	UNHom	30.931	31.788	0	10	40	138
Robberies	Rob	1337.33	2202.379	4	91	1799	14045

Table 3: List of considered loan-specific predictive variables, controls, and corresponding descriptive statistics after having performed the stratified sampling procedure to Banking, Commercial, Financing, Telecommunications, and Utilities as described in Section 5 .

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
Professionals	Prof.	0.006	0.076	0			1
Recovery Rate	RR	0.344	0.43	0	0	0.91	1
Total 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

individual credits available for sampling.

These portfolios can be representative of a third-party collector that, at the end of each quarter, forms portfolios using 100 of the defaulted consumer credits acquired during the reference quarter. 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 predictor set for the portfolios (contract-specific and macroeconomic, financial, and social predictors) is obtained by the weighted averages of the corresponding predictors using the same weights.

3.4 Inter-temporal and out of sample evaluation

The literature on recovery rates modeling has mostly focused on in-sample vs out-of-sample frameworks to analyze recovery rates determinants (Gambetti et al., 2019) or discuss the best predictive strategies for bonds and loans recovery rates (Bellotti et al., 2021; Nazemi et al., 2018). Nevertheless, Kalotay and Altman (2016) and Nazemi et al. (2022) question the applicability of conventional out-of-sample evaluation in the field. Specifically, when the data is randomly partitioned into in-sample and out-of-sample sets without consideration for the time dimension because, then, realizations of recovery rates used for training the models have likely occurred after the defaults of credits used for evaluating the models. The underlying assumption is that the data-generating process is time-invariant, which potentially leads to a look-ahead bias in the predictions. Moreover, in practice, third-party collectors can only use the information available at the time of acquisition of the defaulted credits to estimate their predictive model. This time inconsistency may pose some questions on the statistical appropriateness of standard in-sample vs out-of-sample analysis but also the practical relevance of the results for real-world users. To address this potential issue, we consider as a robustness check an *inter-temporal* prediction exercise by ensuring that only sample points observed before the default event were used during the training process.

4 Forecasting methods

Academic research highlights 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.[6] The promising performance and the flexibility of machine learning, however, come at some costs. First, the estimated models are not always easy to interpret and it is not clear whether they also capture well-known economic and financial relationships. Second, it is not easy to identify the *right tool for the job* in the plethora of supervised ML algorithms. Third, machine learning methods can be computationally expensive. This last point is of particular concern in our study which counts millions of individual credits.

In this study, we employ 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

[6] For example, given that recovery rates are defined in the closed interval $[0, 1]$, predictions arising from a linear regression framework may lead to negative LGDs or values exceeding the unity, also invalidating inference based on Gaussian errors. Moreover, there is empirical evidence that the distribution of individual credit recovery rates is bi-modal, with no or full recovery as the most likely outcome of the collection process. As a result, the relationship between recovery rates and their determinants becomes inherently nonlinear.

splines. We also consider 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). Table 4 reports the predictive algorithms and the forecast combinations employed in this study that we now proceed to briefly describe.

Table 4: List of considered algorithms and corresponding `R` software.

Description	Acronym	`R` algorithm	Reference
Equally weighted average forecast	EW FC		
Optimal FC(+)	Opt_+	`quadprog`	
Optimal FC(+) with shrinkage to EW	COSE_+ FC	`quadprog`	Roccazzella et al. (2022)
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. The shrinkage estimator of the covariance matrix of forecasting error is estimated using `covEstimation` (Ardia et al., 2017)[a] and `FinCovRegularization` (Yan and Lin, 2016)[b] `R` packages. We refer to the Appendix for `R` pseudo-code.

[a] https://CRAN.R-project.org/package=RiskPortfolios.
[b] https://CRAN.R-project.org/package=FinCovRegularization.

4.1 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 $\boldsymbol{y}$ and the $n \times p$ matrix of predictive variables $\boldsymbol{X}$ in a beta regression model, which is tailored to model variables in the interval $(0, 1)$.[7] Following Ferrari and Cribari-Neto (2004), the beta regression model assumes that the dependent variable beta-distributed, with density

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

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

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

$$\mu_i = \frac{e^{x_i^T \boldsymbol{\beta}}}{1 + e^{x_i^T \boldsymbol{\beta}}} \quad \text{and} \quad \phi_i = \phi \; . \tag{2}$$

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,

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u} \; . \tag{3}$$

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.

4.2 Gradient boosted regression trees and random forests

Tree-based methods can identify clusters of data with similar properties, to capture non-linear relationships between the target and its predictive variables as well as to reproduce step-like relationships. Ensembles combine weak learners[8] 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 trees[9] 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 randomly selecting only a subset of predictors at each split of random forests. In this work, we consider an ensemble of 5,000 trees[10] 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

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

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

[10] Following 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.

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 data set 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.

4.3 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:

$$(x_j - t_j)_+ = \begin{cases} x_j - t_j, & \text{if } x_j > t_j \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad (t_j - x_j)_+ = \begin{cases} t_j - x_j, & \text{if } x_j > t_j \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $j = (1, 2, ..., p)$ and $t_j \in \{x_{1j}, x_{2j}, ..., x_{nj}\}$. 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.

4.4 Combinations of forecasts

We now describe methods to aggregate m individual forecasts into a single one. Formally, given a sample of n observations, let $\widehat{\boldsymbol{Y}} = \{\widehat{\boldsymbol{y}}_i, i = 1, ..., m\}$ be the $n \times m$ matrix containing the n forecasts produced by the m models, the vector of the aggregate forecasts is $\widehat{\boldsymbol{y}}_{FC} := \Phi(\widehat{\boldsymbol{Y}})$ using some aggregating function Φ: $\mathbb{R}^{n \times m} \to \mathbb{R}^n$.

In the special case of linear combinations, $\Phi(\widehat{\boldsymbol{Y}})$ takes the form of a weighted average $\widehat{\boldsymbol{Y}}\boldsymbol{w}$ where the weights w_i can be fixed (e.g., to $w_i = \frac{1}{m}$, leading to the simple average forecast) or estimated. A standard choice is to minimize the variance of the aggregate forecast error. In this case, assuming that the individual forecasts are unbiased, restricting the weights to sum to one keeps the aggregated forecast unbiased as well (Timmermann, 2006). With the sum-to-one and nonnegativity constraints on the combining weights, the optimization problem becomes

$$\boldsymbol{w}^* := \underset{\boldsymbol{w} \,\in\, \mathcal{W}}{\operatorname{argmin}} \, \mathbb{E}\left\|\boldsymbol{y} - \widehat{\boldsymbol{Y}}\boldsymbol{w}\right\|_2^2 \quad \text{where} \quad \mathcal{W} := \{\boldsymbol{w} \in \mathbb{R}^m \mid \boldsymbol{1}^T\boldsymbol{w} \; = \; 1, \min \boldsymbol{w} \geq 0\} \,, \tag{5}$$

or, equivalently (Granger and Ramanathan, 1984):

$$\boldsymbol{w}^* := \underset{\boldsymbol{w} \,\in\, \mathcal{W}}{\operatorname{argmin}} \, \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w} \quad \text{where} \quad \mathcal{W} := \{\boldsymbol{w} \in \mathbb{R}^m \mid \boldsymbol{1}^T\boldsymbol{w} \; = \; 1, \min \boldsymbol{w} \geq 0\} \,, \tag{6}$$

where $\boldsymbol{\Sigma}$ is the population covariance matrix of models' prediction error, $\boldsymbol{v}_i := \boldsymbol{y} - \widehat{\boldsymbol{y}}_i$. This problem can be solved efficiently using quadratic programming or imposing the Karush-Kahn-Tucker conditions. Restricting the weights to be non-negative and to sum to one implicitly selects which forecasts to combine and also reduces the estimation error by inducing a shrinkage-like effect on the covariance matrix of forecast errors (Jagannathan and Ma, 2003).

In practice, of course, the population covariance matrix of the models' prediction error is unknown. Therefore, we rely on the plug-in estimator and replace $\boldsymbol{\Sigma}$ by its sample counterpart $\boldsymbol{S}$ in the solution vector.[11]

We also explore the relevance of nonlinear weighting schemes for the forecast combination. Specifically, we opt for averaged neural networks (NN_{FC}) 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. NN_{FC} is an ensemble of one-hidden-layer feed-forward neural networks (Ripley, 1996) and with the *sigmoid* activation function. The

[11] As a robustness check, we also consider the shrinkage estimator towards the diagonal matrix $\boldsymbol{S}_{shrink}$, which is equivalent to shrinking the optimal forecast towards the simple average prediction (Roccazzella et al., 2022). Results are comparable and available upon request.

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.[12] These features moderate the risk of over-fitting.

5 Methodology

5.1 Data sampling for individual NPCs and portfolios

We perform our analysis at individual credit level using training and testing samples that are drawn without replacement using stratified sampling based on the origin of the defaulted credit, i.e., whether it relates to the telecommunication, utilities, banking, or consumer finance industry. This procedure balances the sample for the type of defaulted consumer credit and it is equivalent to the stratified sampling procedure based on seniority types commonly used for defaulted corporate bonds in Nazemi et al. (2017, 2018) and Gambetti et al. (2022). Therefore, the actual number of observations is reduced to 1,204,755 individual credits equally spread between the considered industries. We consider a 70% - 30% split between the training and testing sets.

We now describe how we formed the portfolios of defaulted consumer credits we use for our analysis. For each quarter between (and including) 2007 Q1 and 2019 Q4, we randomly draw (without replacement) samples of 100 individual credits that have been acquired in that quarter. For each quarter, we continue to form portfolios till we have less than 100 individual credits available for sampling. We consider portfolios of 100 credits because in some quarters we have at most 200 credits. This also 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 8 displays the number of portfolios over time.

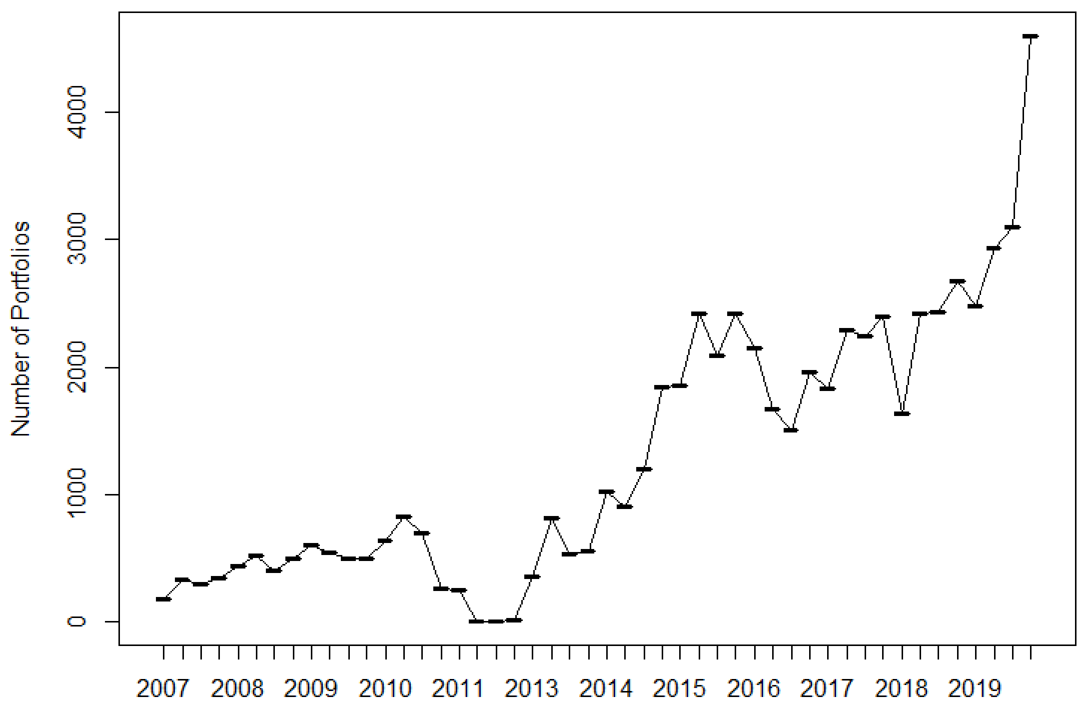
In total, we construct 62,082 portfolios and we consider a 70% - 30% split between training and testing sets. This implies that 43,457 observations are used for training while 18,625 were for the out-of-sample evaluation.

5.2 Time consistency: the inter-temporal exercise

Section 5.1 assumed that the data-generating process is time-invariant when randomly sampling data into in-sample and out-of-sample sets without consideration for the time

[12]Coefficients (called weights in neural networks) are prevented to become 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.

Figure 8: Number of portfolios over time.



dimension. Alternatively, we also design an *inter-temporal* prediction exercise which ensures that the training process only features samples sample points preceding the default event. Specifically, we use NPCs acquired before January 2017 for training, while credits acquired after January 2017 were used for testing. This corresponds to an approximately 50% - 50% split between training and inter-temporal testing sets permitting us to have enough observations to train data-intensive predictive models while keeping enough variability of yearly MS series.[13]

5.3 Evaluation criteria

We now describe the evaluation criteria to measure the potential benefit of including macroeconomic and social indicators in the predictors set while highlighting the differences (if any) between predicting recovery rates at the portfolio and individual credit levels.

[13] As a robustness check, we 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 also robust to this specification and they are available upon request.

We consider the root mean square error (RMSE), the mean absolute error (MAE), and the R^2. For a forecasting strategy j,

$$RMSE_j = \sqrt{\frac{1}{n_{test}} \sum_{i}^{n_{test}} (y_i - \widehat{y}_{i,j})^2} \; ,$$

$$MAE_j = \frac{1}{n_{test}} \sum_{i}^{n_{test}} |y_i - \widehat{y}_{i,j}| \; ,$$

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

where n_{test} is the number of observations in the test set, y_i is the i^{th} realization of the recovery rate, $\widehat{y}_{i,j}$ is the i^{th} forecast of the recovery rate obtained following the strategy j and $\overline{y}$ is the average recovery rate in the training set. Lower values of RMSE and MAE are preferred. Notice that R^2 can be negative.[14] This signals that the forecast method j performs worse than a strategy considering the average recovery rate in the training set as a forecast.

Following Nazemi et al. (2022), we also include the Theil Inequality Coefficient (TIC), which is defined as

$$TIC = \frac{\sqrt{\frac{1}{n_{test}} \sum_i \left(y_i - \hat{y}_{i,j}\right)^2}}{\sqrt{\frac{1}{n_{test}} \sum_i y_i^2} + \sqrt{\frac{1}{n_{test}} \sum_i \hat{y}_{i,j}^2}} \; .$$

TIC rearranges the mean square error as the sum of the average quadratic realized and the estimated recovery rate. This indicator is constrained in $[0, 1]$ and a value closer to 0 indicates higher prediction accuracy.

We compute the averages of out-of-sample (and inter-temporal) performance measures for the models via bootstrap, considering samples of 100 observations and repeating the procedure 10,000 times. In the inter-temporal exercise, observations are bootstrapped at the quarterly level to handle the different number of portfolios over time. As in Nazemi et al. (2022), we use the Mann–Whitney test to assess whether the differences when including (*baseline specification*) and excluding macroeconomic financial and social (*exc. MS specification*) indicators predictors are statistically significant.

[14]Notice that we focus on the *standard* R^2 instead of the *adjusted* R^2. Indeed, while the use of the adjusted R^2 could be motivated by the use of in-sample data to assess the goodness-of-fit independently from the number of predictive variables, we overcome such mechanical bias due to in-sample overfitting by only focusing on out-of-sample (or intertemporal) performance, with observations that have not been used for estimating the forecasting models. Moreover, because of the large size of the testing set relative to the number of predictive variables, this is likely to be irrelevant. For instance, we count 18,625 observations in the testing set at the portfolio level and 43 (or 30) predictive variables when including (or excluding) MS predictors.

To identify the best forecasting approach(es) overall, we consider the model confidence set (hereafter MCS, Hansen et al., 2011) with respect to the square loss function, 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 with significance level α.[15]

6 Results

Next, we present our findings. First, we examine the performance of the out-of-sample forecasting exercise on individual NPCs. Then, we proceed to the out-of-sample and intertemporal exercise on portfolios of NPCs. When dealing with portfolios, all forecasting models are re-estimated using predictive variables at the pool level, mimicking a scenario in which third-party collectors lack information about the particular components of the pool and only possess data at the aggregate portfolio level. Finally, we employ Accumulated Local Effects (ALE) plots (Apley and Zhu, 2020) to depict how individual explanatory variables impact the model predictions, on average. This enables us to determine whether linear and beta regressions, as well as machine learning predictive models, also capture theoretically sound relationships between recovery rates and predictive variables.

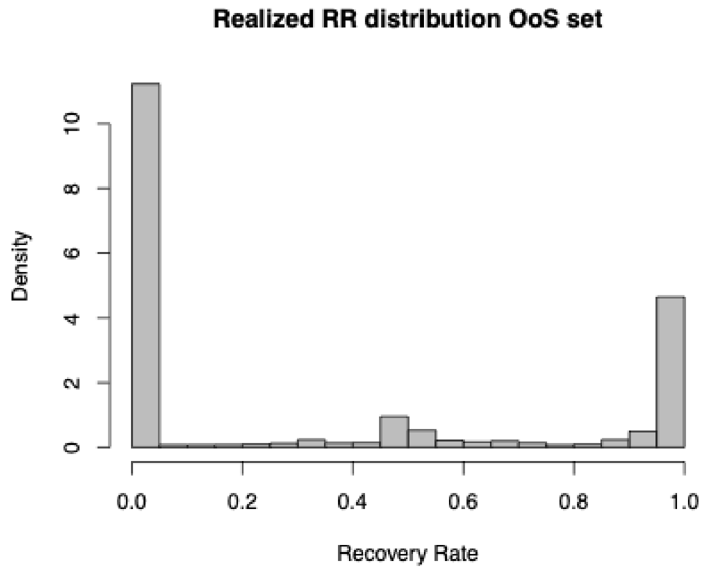
Table 5 reports the RMSE, the MAE, the R^2, the TIC, and the models joining the superior set of models with 5% significance level. Table 6 reports the percentage difference of the performance when excluding MS indicators from the predictor set and the corresponding significance level of the Mann–Whitney test.

6.1 Forecasting performance

Our attention now turns to the outcomes of our predictive exercise using the standard out-of-sample setup. As shown in panels a) and b) of Table 5, we observe a substantial enhancement in predictive accuracy when working with portfolios, regardless of whether we incorporated MS predictors or not. For instance, if we choose Opt+ FC when MS predictors are excluded, the R^2 value rises from 28.27% when examining individual loans to 32.4% in the context of portfolios. This is not surprising because, similarly to the aggregation of individual stocks into portfolios, the idiosyncratic risk of individual credits' recovery rates tends to be diversified away. Moreover, the distribution of realized recovery rates for individual credits strongly differs from the one in the case of portfolios.

[15]Formally, let $\mathcal{M}_0$ be the set of all forecasting models (both individual candidates and forecasts combinations), and let $\mathcal{M}^*$ be the superior set of models. Formally, the MCS tests $H_0 : \mathbb{E}\left[d_{i,j}\right] = 0,\ \forall i, j$. If the null hypothesis is rejected, then the procedure eliminates the model with the greatest relative loss from the set $\mathcal{M}_0$. This procedure is sequentially repeated until the null hypothesis is not rejected at the chosen probability level α. For the MCS, we consider the square loss and compute the p-values via bootstrapping (5,000 replications).

Figure 9: **Distributions of the realized recovery rate for defaulted individual credits and portfolios of consumer credits .**



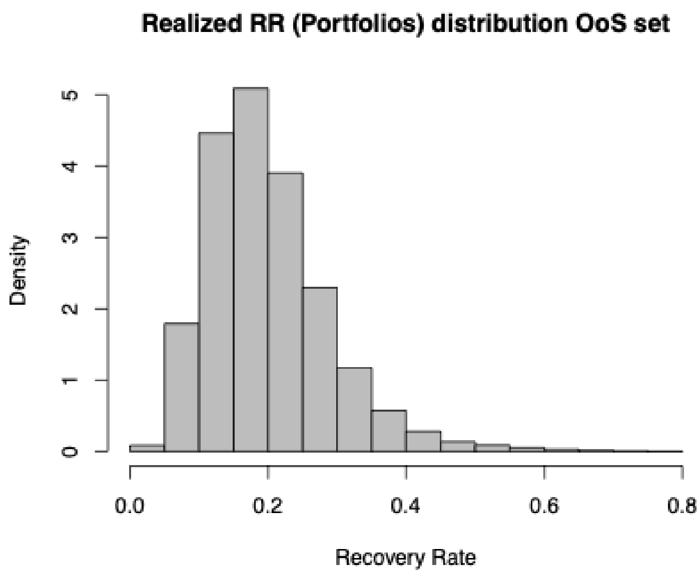


Figure 9 displays how the aggregation into portfolios affects the recovery rates distributions: the cases of no or full recovery (0 and 100%) are diluted in portfolios, offsetting the bi-modality typical of individual recovery rates distributions. This observation is consistent with the theoretical result that describes the asymptotic distribution of the averaged recovery rate on an equally-weighted infinitely granular pool, where defaults are controlled by a one-factor Gaussian copula (Appendix 8.4).

As an additional finding, Table 5 also suggests that forecast combination techniques provide the most effective approach for predicting recovery rates at the individual credit and portfolio levels. In particular, the use of artificial neural networks as a nonlinear combination strategy is particularly encouraging since it is consistently included in the superior group of models and it attains the lowest MAE and RMSE values when MS indicators are included and when dealing with individual credits, portfolios, and portfolios in the inter-temporal setting.

Our focus now shifts to the importance of macroeconomic and social indicators. Panels a) and b) of Table 6 display the percentage differences in performance measures when MS indicators are excluded from the predictor set in the out-of-sample exercise for individual credits and portfolios, respectively. Performance significantly deteriorates when macroeconomic and social indicators are eliminated from the analysis. This is evident when forecasting the recovery rates of defaulted individual credits (panel a), but it is even more apparent at the portfolio level (panel b). When examining individual loans, the maximum improvement in R^2 is approximately 3 percentage points, achieved by using XGBt. However, the maximum improvement in portfolios is more than twice that amount: including MS predictors assists RF in increasing its R^2 by more than 6 percentage points.

Next, we perform the inter-temporal forecasting exercise that explicitly considers the time dimension in the estimation and testing steps. Panel c) of Table 5 indicates that forecasting performance deteriorates compared to the out-of-sample exercise, supporting the findings of Kalotay and Altman (2016) and Nazemi et al. (2022). However, Table 6 also shows how relevant and positive the effect of including macroeconomic and social indicators is. Specifically, all models improve their TIC; nine out of ten models improve their RMSE and R^2, and eight out of ten also 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 linear regression models exhibit an increase of more than 15 percentage points.

6.2 Opening the black box using ALE plots

We have previously found that excluding macroeconomic and social indicators harms forecasting performance. Now, we focus on how economic as well as contract-specific variables influence the prediction of the models on average. To achieve this, we utilize Accumulated Local Effects (ALE) plots (Apley and Zhu, 2020), which show the changes in model predictions within a small neighborhood of the predictor of interest for data instances in that neighborhood.[16] We re-estimate the models by extending the training set to the full sample at the portfolio level to more precisely estimate ALEs for the entire range of values of the considered predictors. We selected the ALE plots in Figure 10, Figure 11, and Figure 12 based on their interpretability and the economic significance of the effect. The vertical axes measure the differences for the average prediction, while the horizontal axes the domain of the corresponding variable.

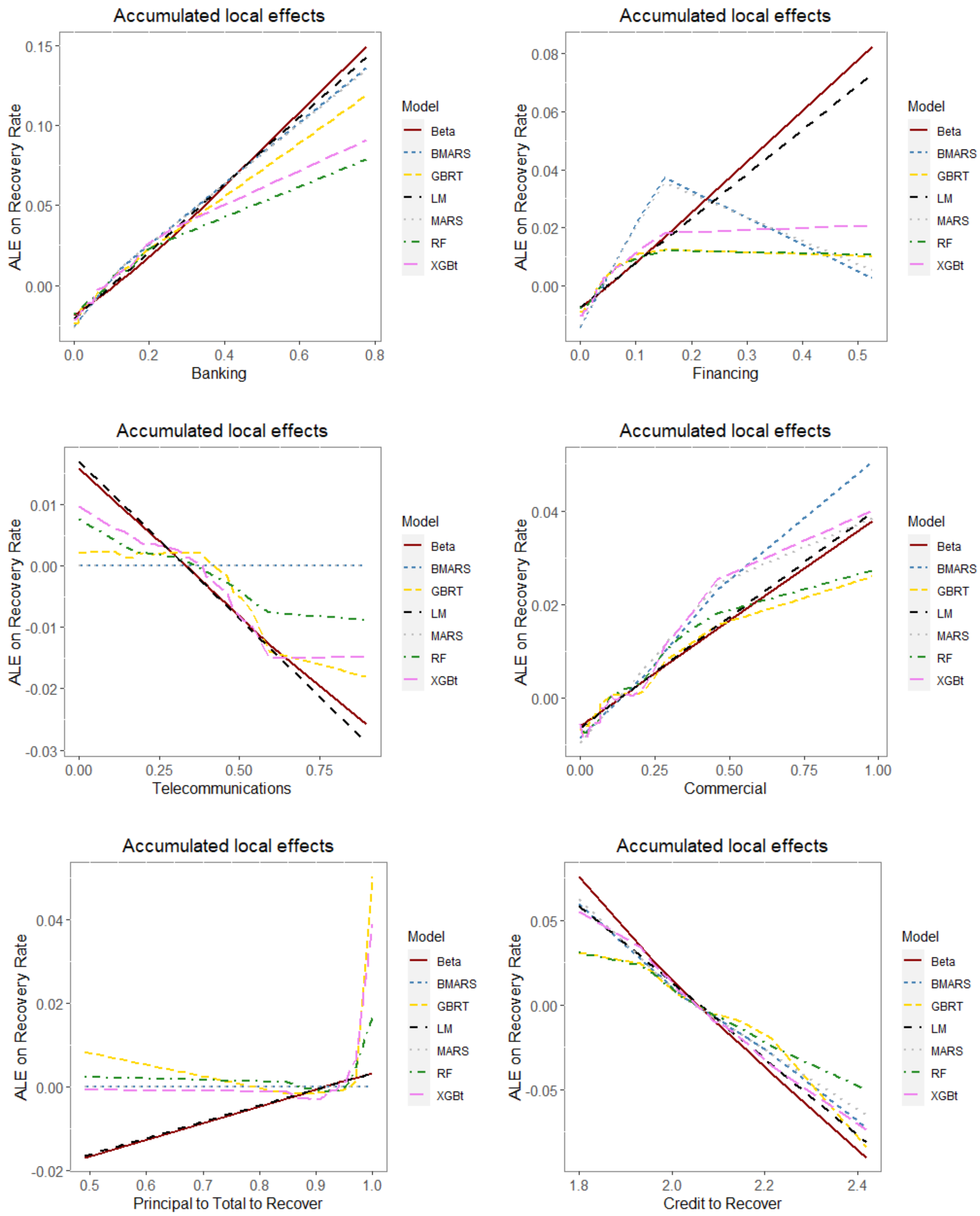
We start studying whether the industry originating the credit affects average recovery rates. ALE plots confirm what we have already observed in the summary statistics in Table 1. Setting the utility sector as a baseline, in *Banking* (Figure 10), we observe that models predict greater recovery rates when the proportion of credits from the banking sector in the portfolio increases. We observe a similar but less accentuated effect related to the quota of credits from the financing industry (*Financing*) and commercial credits (*Commercial*), but, conversely, the greater the proportion of defaulted credits from the telecommunication sector in the portfolio, the lower the expected recovery rate. This finding questions whether studies examining recovery rates for a specific sector can be generalized in the consumer credit industry, as previous research suggests. For instance, the greater expected recovery rate in *Banking* may reflect the better screening ability of credit applicants by banking institutions or could suggest a more efficient recovery process for such types of loans, e.g., because defaulted debtors may fear being banned from the banking system if failing to repay a bank loan.

In the case of *Credit to Recover* (Figure 10), we confirm the negative correlation between the (log of the) total amount to recover and the recovery rate, as previously identified in the summary statistics presented in Table 1. This negative relationship has also been observed in defaulted consumer credits in the telecommunications industry (Nazemi et al., 2022) and the corporate bond market (Bellotti et al., 2021). Additionally, *Principal to Total to Recover* (Figure 10) indicates that a greater proportion of interest and ancillary fees in the total amount to recover leads to a lower expected recovery rate.

Moving to Figure 11, we observe that having been granted more time to perform the collection does not imply greater expected recovery rates (*Max Time to Recover*). Indeed, lenders can accept granting the third-party collector more time for the collection process if they expect the credit to recover to be more challenging. Setting the Lazio region as a baseline, the region where the debtor resides may still affect expected recovery rates, showing that there are still many region-specific factors explaining recovery rates beyond the

[16] Another visualization technique is Partial Dependence (PD) plots, but they tend to be more computationally expensive and may produce inaccurate results when the predictors are highly correlated, as they require extrapolation of the prediction to values of the explanatory variable that are well outside the multivariate envelope of the training data. For a more in-depth discussion of ALE, PD, and marginal plots, we refer readers to Apley and Zhu (2020).

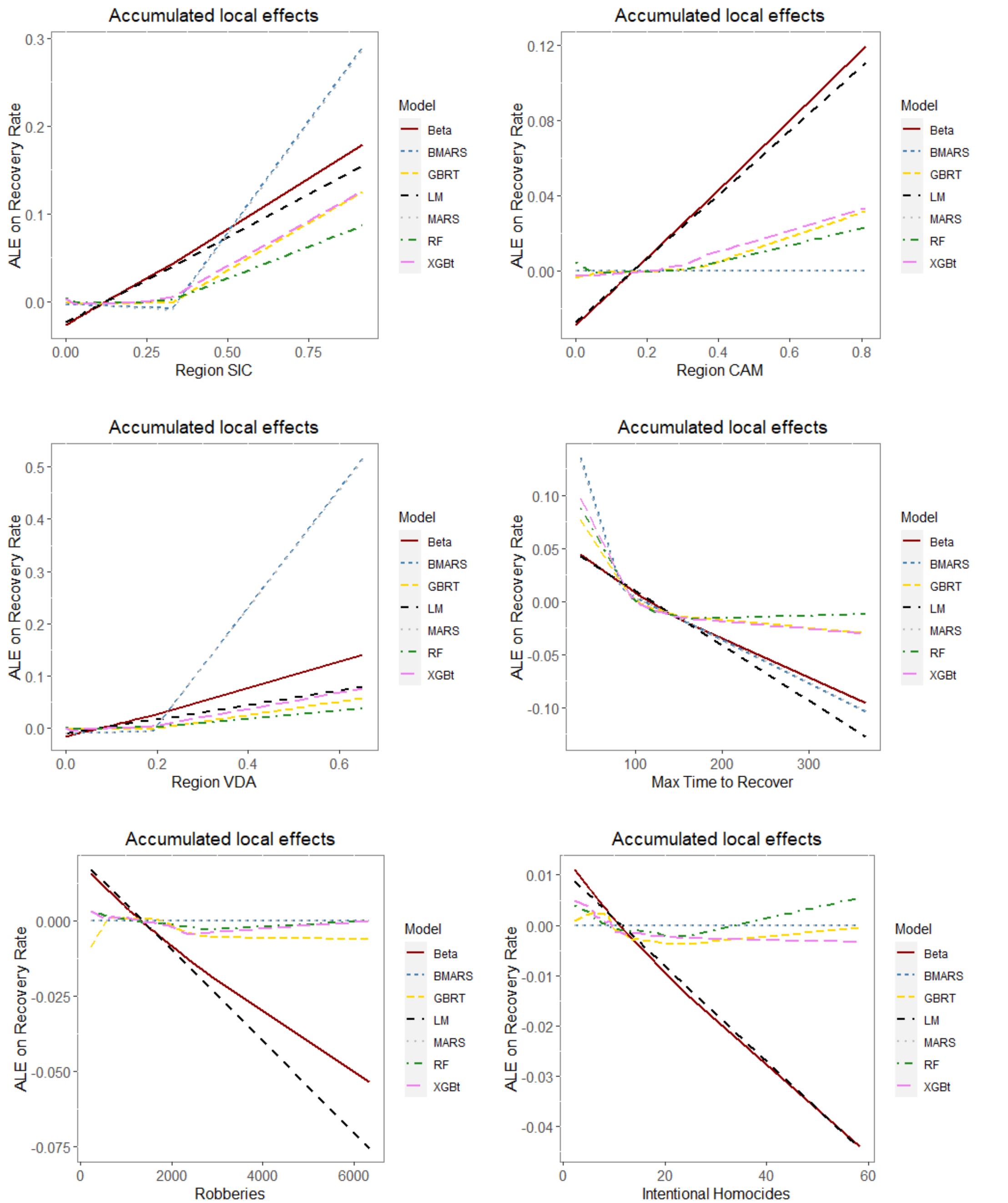
Figure 10: **Accumulate Local Effect Plots.**



ones included in this study. Linear and beta regressions also capture a negative relationship between our proxies of criminal activity, i.e., the number of robberies and intentional homicides per 100.000 inhabitants, and recovery rates. However, the same variables seem to be irrelevant when looking at machine learning techniques.

In Figure 12, *Cost of Debt to GDP* helps us to study whether the dynamics of the credit

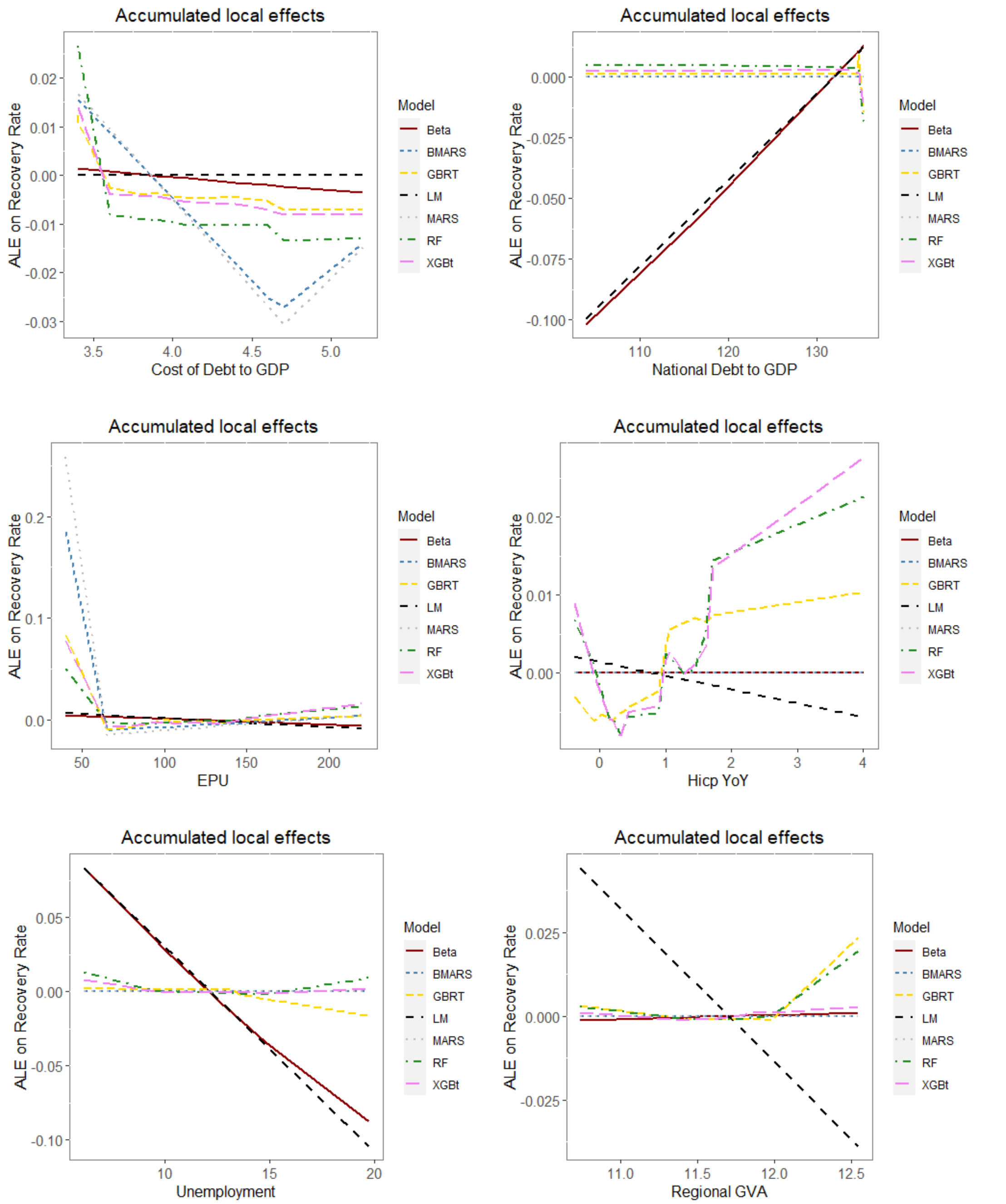
Figure 11: **Accumulate Local Effect Plots.**



cycle at the national level influence recovery rates for consumer credits. *Cost of Debt to GDP* signals how a higher level of interest to GDP at the Italian level can be associated with lower expected recovery rates. Similarly, we notice that only when the Italian debt to GDP ratio goes above 135%, recovery rates are expected to slightly decrease.

Figure 12 allows us to examine whether the national-level credit cycle dynamics affect

Figure 12: **Accumulate Local Effect Plots.**



recovery rates for consumer credit by focusing on *Cost of Debt to GDP*. We observe that higher levels of public debt interest in Italy are associated with lower expected recovery rates. Additionally, we notice that recovery rates are expected to slightly decrease when the *National Debt to GDP* ratio exceeds 135%. When looking at *EPU*, i.e., the ALE plot associated with the Economic Policy Uncertainty index of Baker et al. (2016), we highlight how low uncertainty can be associated with higher expected recovery rates. Among the MS variables, the *EPU* has the largest effect on expected recovery rates, extending the findings of Gambetti et al. (2019) to defaulted consumer credits.

The accumulated local effect associated with linear and beta regression models underlines the negative relationship between *Unemployment* and expected recovery rate, however, machine learning algorithms seem not to be too sensitive to this predictor. Inflation at the national level, i.e., *HICP YoY*, seems positively associated with higher expected recovery rates, however, the estimated accumulated local effect varies across models. Concerning *HICP YoY*, however, we remark that Italy suffered from a deflationary trend following the outburst of the European sovereign bonds crisis in 2012. Therefore, it is not surprising that higher levels of inflation are also related to higher expected recovery rates. Nevertheless, we underline that the inflation level never went above 4% in the sample under analysis and therefore, we cannot assess the effect of high inflation rates on the recovery rate. Finally, a high level of *Regional GVA*, i.e., real output at the regional level, is also associated with a greater expected recovery rate, but the relationship is highly nonlinear, the size of the effect is only modest and it is only captured by gradient boosted trees.

7 Conclusions

Forecasting the recovery rate is crucial for all stakeholders involved in estimating the actual loss of non-performing consumer credits. Contrary to previous studies that suggested macroeconomic variables had no empirical significance when included in the predictors set, our findings demonstrate the opposite. Exploiting a private database of unprecedented size, we show that not only contract-specific variables are important: macroeconomic predictors boost the prediction performance to a large extent.

Specifically, we find that including macroeconomic and social indicators improves RR forecasting performance both at the individual credit and (especially) portfolio level for credits originating from the utilities, banking, financial and commercial sectors. Although analyzing the determinants at the portfolio level is crucial to closely match the business model of third-party collectors that perform the recovery process in this industry, this was overlooked in earlier studies.

This result is evident when looking at the sharp and significant decrease in forecasting performance of linear and beta regression models but also more advanced, yet standard, machine learning and forecast combination methods. 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 also rather 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, it is worth underlying that the significant improvement in forecasting performance is due to the inclusion of only 13 publicly available macroeconomic and social variables. This is particularly valuable for credit institutions, third-party debt collection firms, or regulatory bodies in charge of capital requirements policies as parsimonious forecasting frameworks are preferred to large predictor sets that are instead potentially overlooked because of access, maintenance, and validation difficulties.

References

Acharya, V. V., S. T. Bharath, and A. Srinivasan (2007, sep). Does industry-wide distress affect defaulted firms? evidence from creditor recoveries. *Journal of Financial Economics 85*(3), 787–821.

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.

Ardia, D., K. Boudt, and J.-P. Gagnon-Fleury (2017). RiskPortfolios: Computation of risk-based portfolios in R. *Journal of Open Source Software 10*(2).

Atiya, A. F. (2020, jan). 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, jul). Measuring economic policy uncertainty. *The Quarterly Journal of Economics 131*(4), 1593–1636.

Bank of Italy (2020). L'economia delle regioni italiane. dinamiche recenti e aspetti strutturali. Technical report.

Beck, T., J. Grunert, W. Neus, and A. Walter (2017, apr). What determines collection rates of debt collection agencies? *Financial Review 52*(2), 259–279.

Bellotti, A., D. Brigo, P. Gambetti, and F. Vrins (2021, jan). 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.

Bruche, M. and C. González-Aguado (2010). Recovery rates, default probabilities and the credit cycle. *Journal of Banking & Finance 34*(4), 754–764.

Chen, T. and C. Guestrin (2016, aug). XGBoost. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM.

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, 03). Multivariate adaptive regression splines. *The Annals of Statistics 19*(1), 1–67.

Friedman, J. H. (2002, feb). Stochastic gradient boosting. *Computational Statistics & Analysis 38*(4), 367–378.

Gambetti, P., G. Gauthier, and F. Vrins (2019). Recovery rates: Uncertainty certainly matters. *Journal of Banking & Finance 106*, 371–383.

Gambetti, P., F. Roccazzella, and F. Vrins (2022). Meta-learning approaches for recovery rate prediction. *Risks 10*(6).

Gordy, M. B. (2003, 07). 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, April). 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, jul). Loss functions for loss given default model comparison. *European Journal of Operational Research 268*(1), 348–360.

Jagannathan, R. and T. Ma (2003, July). 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, jun). 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.

Livshits, I. (2015, aug). Recent developments in consumer credit and default literature. *Journal of Economic Surveys 29*(4), 594–613.

Loterman, G., I. Brown, D. Martens, C. Mues, and B. Baesens (2012, jan). Benchmarking regression algorithms for loss given default modeling. *International Journal of Forecasting 28*(1), 161–170.

Merton, R. C. (1974, may). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance 29*(2), 449.

Nakajima, M. and J.-V. Ríos-Rull (2014, October). Credit, bankruptcy, and aggregate fluctuations. Working Paper 20617.

Nazemi, A., F. Baumann, and F. J. Fabozzi (2022, 03). Intertemporal defaulted bond recoveries prediction via machine learning. *European Journal of Operational Research 297*(3), 1162–1177.

Nazemi, A. and F. J. Fabozzi (2018, apr). Macroeconomic variable selection for creditor recovery rates. *Journal of Banking & Finance 89*, 14–25.

Nazemi, A., K. Heidenreich, and F. J. Fabozzi (2018, dec). 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, oct). 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, jan). To tune or not to tune the number of trees in random forest. *J. Mach. Learn. Res. 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. Vrins (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.

Shleifer, A. and R. Vishny (1992). Liquidation values and debt capacity: a market equilibrium approach. *The Journal of Finance 47*(4), 1343–1366.

Smithson, M. and J. Verkuilen (2006). A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods 11*, 54–71.

Thomas, L., A. Matuszyk, and A. Moore (2012). Comparing debt characteristics and lgd models for different collections policies. *International Journal of Forecasting 28*(1), 196–203.

Timmermann, A. (2006). Chapter 4 forecast combinations. In *Handbook of Economic Forecasting*, pp. 135–196. Elsevier.

Yan, Y. and F. Lin (2016). *FinCovRegularization: Covariance Matrix Estimation and Regularization for Finance.* R package version 1.1.0.

Yao, X., J. Crook, and G. Andreeva (2017, 12). Enhancing two-stage modelling methodology for loss given default with support vector machines. *European Journal of Operational Research 263*(2), 679–689.

Table 5: Performance for the out-of-sample forecasting exercise at individual consumer loan and portfolio level for both specifications of the predictors set, i.e., when including or excluding MS. We highlight in bold the best performance measures for each specification. We report the models that have been included in the superior set of models at a 5% significance level in both MCS - incl. MS and MCS - excl. MS, based on whether they were estimated with or without MS predictors.

	RMSE	MAE	R2	TIC	RMSE	MAE	R2	TIC
	Predictors set excluding MS				Predictors set including MS			
a) Out of sample - Individual loans								
MARS	.3840	.3213	.1980	.4082	.3814	.3176	.2089	.4045
GBRT	.3676	.3004	.2646	.3864	.3632	.2952	.2818	.3809
XGBt	.3736	.3040	.2396	.3811	.3661	.2868	.2695	.3717
RF	.3673	**.2915**	.2654	.3805	.3615	.2867	.2880	.3730
B-MARS	.3847	.3249	.1952	.4112	.3822	.3210	.2054	.4071
Beta	.4282	.3990	.0046	.4576	.4247	.3946	.0205	.4528
OLS	.4027	.3473	.1189	.4370	.3987	.3422	.1364	.4306
Opt+ FC	**.3630**	.2960	**.2827**	**.3789**	.3567	.2884	.3074	**.3705**
EW FC	.3734	.3241	.2422	.3997	.3684	.3180	.2624	.3931
ANN FC	.3633	.2935	.2814	.3816	**.3561**	**.2801**	**.3093**	.3712
MCS L2 - incl. MS	ANN FC							
MCS L2 - excl. MS	·							
b) Out of sample - Portfolios								
MARS	.0745	.0556	.2820	.1759	.0725	.0540	.3198	.1708
GBRT	.0733	.0548	.3050	.1728	.0705	.0528	.3570	.1659
XGBt	.0740	.0559	.2912	.1726	.0718	.0542	.3314	.1672
RF	.0734	.0553	.3026	.1727	.0700	.0529	.3652	.1645
B-MARS	.0738	.0552	.2974	.1742	.0736	.0548	.3022	.1737
Beta	.0790	.0583	.2013	.1869	.0768	.0567	.2438	.1815
OLS	.0790	.0586	.1998	.1872	.0768	.0571	.2421	.1818
EW FC	.0733	.0551	.3078	.1731	.0711	.0535	.3489	.1676
Opt+ FC	**.0723**	**.0545**	.3240	.1703	**.0697**	.0526	**.3715**	.1639
ANN FC	.0724	.0546	.3205	**.1697**	**.0697**	**.0524**	.3690	**.1629**
MCS L2 - incl. MS	Opt+ FC ; COSE+ FC ; ANN FC							
MCS L2 - excl. MS	·							
c) Intertemporal - Portofolios								
MARS	.0813	.0592	.1590	.1894	.0772	.0589	.2414	.1663
GBRT	.0798	.0584	.1919	.1860	.0760	.0562	.2687	.1706
XGBt	.0852	.0636	.0802	.2053	.0809	.0607	.1712	.1884
RF	**.0780**	**.0576**	**.2280**	**.1813**	.0785	.0617	.2153	.1690
B-MARS	.0820	.0601	.1454	.1940	.0767	.0577	.2524	.1683
Beta	.0848	.0632	.0891	.2032	.0774	.0582	.2396	.1754
OLS	.0868	.0656	.0434	.2090	.0788	.0598	.2119	.1787
EW FC	.0811	.0601	.1665	.1921	.0758	.0572	.2721	.1690
Opt+ FC	.0784	.0578	.2199	.1829	.0764	.0590	.2580	.1671
ANN FC	.0793	.0584	.2015	.1860	**.0753**	**.0572**	**.2796**	**.1651**
MCS L2 - incl. MS	ANN FC							
MCS L2 - excl. MS	·							

Table 6: Loss differentials for the out-of-sample forecasting exercise when excluding macroeconomic, financial, and social indicators.

	RMSE		MAE		R2		TIC	
a) Out-of-sample - Individual loans								
MARS	0.69%		1.17%		-1.08		0.91%	
GBRT	1.20%		1.76%	*	-1.72	*	1.44%	
XGBt	2.06%	**	6.00%	***	-2.99	**	2.52%	**
RF	1.58%		1.67%		-2.26	*	2.01%	*
B-MARS	0.65%		1.21%		-1.02		1.01%	
Beta	0.82%	*	1.13%	***	-1.59	***	1.05%	***
OLS	1.01%	*	1.49%	**	-1.74	**	1.49%	***
EW FC	1.37%	*	1.92%	**	-2.03	**	1.68%	**
Opt+ FC	1.78%	*	2.63%	**	-2.47	**	2.26%	**
ANN FC	2.03%	*	4.80%	***	-2.79	**	2.81%	**
b) Out-of-sample - Portfolios								
MARS	2.76%	**	2.96%	**	-3.78	**	2.99%	*
GBRT	3.97%	**	3.79%	**	-5.20	***	4.16%	**
XGBt	3.06%		3.14%	**	-4.02	*	3.23%	**
RF	4.86%	**	4.54%	***	-6.26	***	4.98%	***
B-MARS	0.27%		0.73%		-0.48		0.29%	
Beta	2.86%	*	2.82%	*	-4.25	***	2.98%	**
OLS	2.86%	*	2.63%	*	-4.23	***	2.97%	**
EW FC	3.09%	*	2.99%	**	-4.11	***	3.28%	**
Opt+ FC	3.73%	*	3.61%	**	-4.75	***	3.90%	**
ANN FC	3.87%	*	4.20%	***	-4.85	**	4.17%	**
c) Intertemporal - Portfolios								
MARS	5.31%	***	0.54%		-8.24	***	13.91%	***
GBRT	5.05%	***	3.93%	**	-7.68	***	9.01%	***
XGBt	5.33%	***	4.77%	***	-9.09	***	8.96%	***
RF	-0.60%		-6.66%	***	1.27		7.27%	***
B-MARS	6.98%	***	4.24%	***	-10.70	***	15.24%	***
Beta	9.52%	***	8.65%	***	-15.05	***	15.86%	***
OLS	10.25%	***	9.78%	***	-16.86	***	16.98%	***
EW FC	7.04%	***	5.19%	***	-10.56	***	13.66%	***
Opt+ FC	2.62%	*	-2.05%	**	-3.80	***	9.44%	***
ANN FC	5.36%	***	2.03%		-7.81	***	12.62%	***

8 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.

8.1 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))
```

8.2 Minimum variance combination of forecasts with non negative weights

Using the R package quadprog[17], the function Weights_P estimates the combining weights given the covariance matrix Qmat.

```
Weights_P <- function(Qmat){
```

[17] https://cran.r-project.org/package=quadprog.

```
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)}
```

CO+ Constrained Optimization.

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

COSE+ Constrained Optimization with Shrinkage toward diagonal matrix[18]

```
S_shrink_D <- covEstimation(as.matrix(FE), control = list(type ='diag'))
Weights_shrink_D  <- Weights_P(_shrink_D)
```

8.3 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))
```

[18] Required packages: `Ardia2017` (Ardia et al., 2017)

8.4 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 $1, \ldots, n$ 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 assumptions[19], the variable $\langle R_n \rangle$ converges to its conditional expectation $\langle R_n(Z) \rangle := \mathbb{E}[\langle R_n \rangle | Z]$ as $n \to \infty$ (see Gordy, 2003, or 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 diversified away perfectly, 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 $\pi = \mathbb{P}(R_i = 1)$. In this illustration, we consider the one-factor Gaussian copula dependence scheme, which is the corner stone of the loss model adopted in the Basel regulation on capital requirements[20] :

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

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

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

$$= \frac{1}{n} \sum_{i=1}^n \mathbb{P}(X_i \leq \Phi^{-1}(\pi) | Z) = \Phi\left(\frac{\Phi^{-1}(\pi) - \sqrt{\rho} Z}{\sqrt{1-\rho}}\right) \ .$$

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

[20] In 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.

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

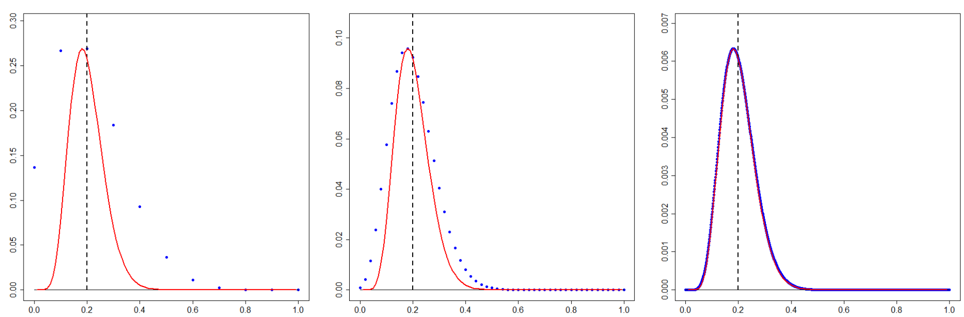
$$\begin{aligned}
\mathbb{P}(\langle R_n(Z) \rangle \leq x) &= \mathbb{P}\left(\Phi\left(\frac{\Phi^{-1}(\pi) - \sqrt{\rho}Z}{\sqrt{1-\rho}}\right) \leq x\right) \\
&= \mathbb{P}\left(Z \geq \frac{\Phi^{-1}(\pi) - \sqrt{1-\rho}\Phi^{-1}(x)}{\sqrt{\rho}}\right) \\
&= \Phi\left(\frac{\sqrt{1-\rho}\Phi^{-1}(x)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

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

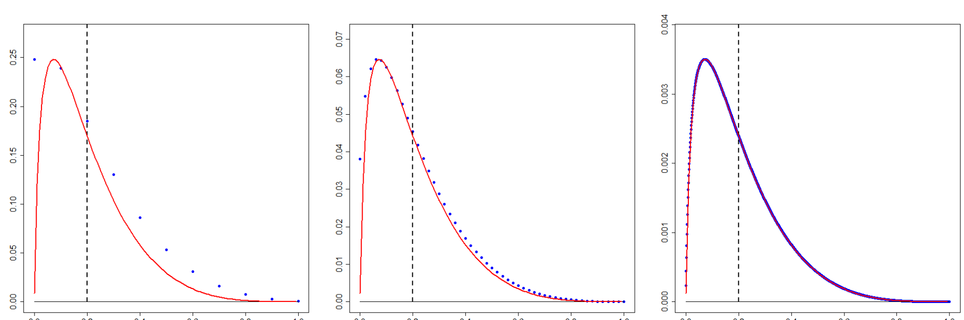
$$\hat{f}_{\langle R_n \rangle}(x) := \frac{1}{\varphi(\Phi^{-1}(x))}\sqrt{\frac{1-\rho}{\rho}}\varphi\left(\frac{\sqrt{1-\rho}\Phi^{-1}(x) - \Phi^{-1}(\pi)}{\sqrt{\rho}}\right) \tag{7}$$

Figure 13 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 $\pi = \mathbb{E}[R_i] = 0.2$ 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. 7 offers quickly a good approximation of the exact distribution associated with the recovery rate model (in practice, as from n is equal to a few tenths).

Figure 13: Exact (blue dots) and asymptotic (red solid line) distributions of the average recovery rate on an equally-weighted homogeneous pool ($w_i = 1/n$, $\pi = 0.2$) 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 $\langle R_n \rangle$.



(a) $\rho = 5\%$



(b) $\rho = 25\%$