BUSINESS CYCLE AND REALIZED LOSSES IN THE CONSUMER CREDIT INDUSTRY

Walter Distaso, Francesco Roccazzella, Frédéric Vrins

REPRINT • 2024 / 14

LFIN

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

Tel (32 10) 47 43 04

Email: lidam-library@uclouvain.be

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

Business cycle and realized losses in the consumer credit industry*

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

[1]Imperial College Business School, United Kingdom

[2]UNIME, Italy

[3]IESEG School of Management, Univ. Lille, CNRS, UMR 9221
- LEM - Lille Economie Management, F-59000 Lille, France

[4]UCLouvain, LIDAM - LFIN, Belgium
December 15, 2024

Abstract

We investigate the determinants of losses given default (LGD) in consumer credit. Utilizing a unique dataset encompassing over 6 million observations of Italian consumer credit over a long time span, we find that macroeconomic and social (MS) variables significantly enhance the forecasting performance at both individual and portfolio levels, improving R^2 by up to 10 percentage points. Our findings are robust across various model specifications. Non-linear forecast combination schemes employing neural networks consistently rank among the top performers in terms of mean absolute error, RMSE, R^2, and model confidence sets in every tested scenario. Notably, every model that belongs to the superior set systematically includes MS variables. The relationship between expected LGD and macro predictors, as revealed by accumulated local effects plots and Shapley values, supports the intuition that lower real activity, a rising cost-of-debt to GDP ratio, and heightened economic uncertainty are associated with higher LGD for consumer credit. Our results on the influence of MS variables complement and slightly differ from those of related papers. These discrepancies can be attributed to the comprehensive nature of our database – spanning broader dimensions in space, time, sectors, and types of consumer credit — the variety of models utilized, and the analyses conducted.

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

---

[*] The authors are grateful to the participants of the XVIII Credit Scoring and Credit Control Conference (Edinburgh, 08/30-09/01, 2023) and the 40th International Conference of the French Finance Association (Lille on 27/05-29/05, 2024) for valuable feedback and discussions. F. Vrins thanks Arun Verna, head of Quantitative Research Solutions at Bloomberg L.P., for inspiring discussions about TreeShap.

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

1 Introduction

Consumer credit refers to the debt that individuals contract to purchase goods and services. This includes the usual loans and credit lines (issued by credit institutions or by sellers directly) as well as liabilities resulting from unscheduled deferred payments.[1] Consumer credit is a rapidly expanding industry globally. As of September 2022, it reached USD 4,700 billion in the US alone and EUR 715 billion in the euro area, in which consumer credit constitutes 10.8% of the total outstanding credit for households.[2] The size of this market makes it crucial for creditors and all stakeholders involved in the recovery process to study the drivers of realized losses in the event of default.

When confronted with a non-performing credit (NPC), a debt holder typically has four main options: accept the loss, initiate a collection procedure with the customer, mandate a specialized firm to collect the debt on their behalf or sell the NPC at a discount to a company that will then attempt to recover the defaulted credit by pursuing the customer. Whether evaluating the relevance of initiating a collection procedure, determining the fees of a third-party debt collection firm, or pricing the NPC for sale or purchase, every stakeholder needs to estimate the likely recovery rate (RR). In the event of a default, the severity of the loss suffered by creditors can be measured using two parameters: the exposure at default (EAD) and the RR (or, equivalently, loss given default, LGD = 1 - RR). The EAD refers to the outstanding amount of the debt at the time of default, while the RR denotes the ratio of the amount that creditors will recover to the EAD, hence, takes value in $[0, 1]$. The EAD is typically known at the time of default. In contrast, the RR is only revealed when closing the debt collection process.

Although a vast literature addresses the RRs of defaulted corporate bonds, less is known about those of non-performing consumer credit. Unfortunately, extrapolating findings related to corporate bonds to consumer credits is challenging, as the two types of debt are not easily comparable. The available information, market participants, and amounts at stake differ significantly, to name just a few differences. Additionally, the empirical distributions of realized RRs exhibit different shapes, further highlighting the disparities.[3] When it comes to identifying the main drivers of RRs, most of the literature has documented the central role of contract-specific variables (such as the seniority or the sector for bonds, and the recovery duration or the principal for consumer credits). Conversely, the impact of macroeconomic and social (MS) indicators on RR is less clear and seems to depend on the type of debt at hand. For instance, it has been known since the seminal paper of Altman et al. (2005) that realized recovery rates of corporate bonds are negatively correlated with the default rate, indicating their sensitivity to business cycle fluctuations. Surprisingly enough, this would not apply to consumer credit, where macroeconomic variables have been shown to be empirically irrelevant (see for example, Nazemi et al., 2022). Different factors might have contributed to these findings. First, it is very difficult to access a meaningful consumer credit dataset: data is typically private, often scarce, and limited to a specific sector of the consumer credit industry,

---

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

[2] Data on credit extended to individuals for household, family, and other personal expenditures, excluding loans secured by real estate for the 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/`.

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

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

Our paper offers a novel perspective on how macroeconomic and social factors influence the recovery rates of non-performing consumer credits and makes several contributions to the existing literature. First, we jointly investigate the factors influencing consumer credit recovery rates from multiple industries, specifically from utilities, banking, financial, and commercial sectors. A database encompassing over 6 million NPCs in Italy, covering the period from 2007 to 2019 makes this comprehensive analysis possible. To the best of our knowledge, the dataset at hand is the most extensive examined in academic literature to date. Second, we perform our analysis on two levels: at the individual NPC level and at the portfolio level. While previous research primarily concentrates on the individual level, it is important to consider the portfolio level as well, for the reasons explained above. Third, we exploit the heterogeneity of economic dynamics across regions and provinces of Italy to understand how the economy impacts LGDs in the consumer credit industry.[4] In contrast to previous studies focusing on specific types of NPC and including a limited set of macroeconomic indicators only at the national level, we find that excluding MS indicators from the set of predictors significantly deteriorates the forecasting performance, both at individual credit and portfolio levels. We show that any model that belongs to the superior set of models at the 95% confidence level includes MS variables. Our findings are robust to the choice of forecasting models; they remain valid when considering standard regressions, machine learning algorithms, and combinations of those. Fourth, we find using artificial neural networks as a nonlinear combination strategy particularly encouraging since it attains the lowest MAE (mean absolute error) and RMSE (root mean square error) values. Fifth, while variable importance metrics and Shapley values also support the relevance of MS indicators, the accumulated local effect (ALE) plots confirm that our results are in line with economic intuition. For instance, a deterioration of the credit conditions at the national level is associated with lower RRs.

Our analysis highlights the importance of macroeconomic and social indicators in enhancing models that forecast RRs for consumer credits. While contract-specific factors are the primary drivers, some macroeconomic indicators like economic uncertainty or credit and business cycles indicators offer greater explanatory power than certain idiosyncratic variables. This demonstrates that MS variables provide complementary information, helping to explain variations in RR across different economic environments. Consequently, they offer crucial insights for all stakeholders in the consumer credit industry.

---

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

2 Literature review

In this section, we review the literature on the factors influencing RRs with a particular focus on the significance of MS predictors.

The empirical evidence on the importance of MS indicators for explaining consumer credit RRs is sparse and largely inconclusive. Results seem to differ depending on the timing of the studies, the category and sector of the NPC, and the type of MS variables taken into account.[5]

Leow et al. (2014) found that macroeconomic variables may improve the predictive performance of mortgage loan RR estimates, but, conversely, not of personal loan RR. They considered a broad set of economic variables (including net lending growth, disposable income growth, GDP growth, net lending growth on dwellings, unemployment rate, saving ratio, interest rate, and the Halifax House Price Index) with data on defaulted mortgages (from 1990 to 2002) and unsecured personal (from 1989 to 1999). Beck et al. (2017) examined the RRs of consumer credits for goods and services in the period 2004-2008 in Germany, and focused on idiosyncratic determinants such as the EAD or prior debtor-specific collection rates, complemented by GDP growth and the regional unemployment rate. The latter variable was found to have a consistently negative effect on the RR. The legal environment also plays an important role. In the credit card industry, for instance, Fedaseyeu (2020) found that consumer protection legislation governing third-party debt collection reduces the number of third-party debt collectors and increases the LGD on delinquent credit card loans.

More recently, Nazemi et al. (2022) considered 65,535 defaulted unsecured consumer credits purchased between 2010 and 2013 from a German telecommunications company. In their case, however, including macro variables such as the provincial unemployment rate and excessive indebtedness provincial rate (the percentage of adults with strong negative credit ratings) did not lead to improved prediction accuracy. Hacht and Zagst (2010) analyse the determinants of recovery rates for defaulted bank loans, including both credit and macroeconomic factors, using data from the Global Credit Data databases, which focus on large borrowers and specialized, collateralized lending. They find that macroeconomic variables have only a minor impact at the sector level. In contrast, our study examines non-performing credits (NPCs) such as defaulted utility or telecommunication contracts, involving smaller amounts and different customer types, like individual households and professionals. Methodologically, Hacht and Zagst (2010) use linear models and a forward-backward selection algorithm, concluding that macroeconomic variables play a minor role. Their subsequent work (Höcht et al., 2022) employs a Markov-Switching model to explain the time series dynamics of bank loan recovery rates at the aggregate level, focusing on the European and US economies. This aggregation limits their analysis to macroeconomic predictors and lacks contract-specific

---

[5] A similar debate holds for non-performing loans: Dermine and de Carvalho (2006) find on a dataset covering the period 1995-2000 in the UK that GDP does not matter, in contrast with Bellotti et al. (2021) who find by analysing a 1988-2015 dataset from a European bank that macro indicators are key. Bellotti and Crook (2012) suggest that the limited impact of macroeconomic factors on LGD observed by Dermine and de Carvalho (2006) may result from their smaller training sample size (374 defaults), which might not have been large enough to reveal a significant relationship between the economy and recovery rates.

information, which can be important for recovery rates.

It is reasonable to assume that the ability of households to repay defaulted debts depends on their current financial situation, with low aggregate consumption and high unemployment rates indicating insufficient income and wealth, which complicates loan repayment and affects recovery values. Calabrese (2012) found systematic links between average recovery rates and macroeconomic indicators such as the interest rate, GDP growth, unemployment, and the aggregate default rate using a comprehensive survey on the loan recovery process of Italian banks. Employing aggregated data, Caselli et al. (2008) found that the time series dynamics of LGD on banking loans to households and SMEs depend on different macro factors. For households, the best regression model relates LGD to the household default rate, unemployment rate, and household consumption. Konecny et al. (2017) showed that delayed macroeconomic effects explain RRs in the consumer credit market using data from the Czech retail banking sector, highlighting the importance of considering lagged effects in addition to contemporaneous ones. Bellotti and Crook (2012), considering 55,000 defaulted credit card accounts in the UK from 1999 to 2005, found that incorporating macroeconomic variables can enhance the forecasting power of recovery rates. Higher interest rates and unemployment levels at the time of default are linked to lower recovery rates, while higher earnings growth may lead to better recoveries. Recent studies also highlight the potential of explainable AI models to identify credit risk drivers; for instance, Bussmann et al. (2020) and Bastos and Matos (2022) demonstrate the benefits of using Shapley values and correlation networks for peer-to-peer lending and corporate bonds, respectively. Furthermore, Kellner et al. (2022) and Nazemi and Fabozzi (2024) show the superior predictive performance and insights provided by interpretable machine learning models for bank loans and defaulted U.S. corporate bonds.

Much of credit risk research focuses on credit scoring, which involves estimating default probabilities and their main drivers, as well as understanding credit cycles. For example, Djeundje and Crook (2018) found that incorporating account-specific effects and macroeconomic variables significantly enhances predictive accuracy for credit card defaults. Similarly, Malik and Thomas (2010) showed that consumer default intensities for banking loans are influenced by business cycle indicators, and Carvalho et al. (2020) confirmed that incorporating macroeconomic information improves the accuracy of models forecasting defaults of non-financial firms in the euro area, with GDP growth notably reducing default probabilities. Foglia (2022) leverage data from the Italian credit industry to determine the impact of macroeconomic determinants on the NPL rate (the percentage of loans in default) in the Italian banking system, using a sample of quarterly data from 2008Q3 to 2020Q4 in a time-series setting. In contrast with this existing literature focusing on modeling default probabilities, our purpose is to investigate the impact of contract-specific and MS variables jointly on the recovery rates of defaulted consumer credits.

3 Data

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

3.1 The non-performing credits database

The NPC database contains 6,493,794 instances of defaulted consumer credits that third-party collectors are entitled to recover within a contractually specified maximum period after receiving the mandate from the original lender. The RR can be defined and measured using various approaches. For instance, the recovery rate of defaulted American bonds is defined in Moody's Default and Recovery Database as the market value of the bond 30 days post default (Mora, 2012) while, in the context of credit cards, RR is often defined as the sum of repayments made over a collection period Δ divided by the outstanding balance at the date of default (Bellotti and Crook, 2012; Konecny et al., 2017, considered $\Delta = 12$ month). Similarly to the case of credit cards, the definition we follow in this work is tailored to our specific data and purposes.

We consider NPCs with a maximum collection period of $\Delta = 365$ days from the receipt of the mandate by the third-party collector. At the end of this mandate period, if the recovery is incomplete (i.e., $RR < 1$ and without a settlement), the original creditor may continue collection efforts, either directly or by assigning a second mandate to a specialized firm. In such cases, the RR may be underestimated, making it unsuitable for direct extrapolation as an estimate of ultimate recovery. As noted by Gürtler and Hibbeln (2013) and Rapisarda and Echeverry (2013), the underestimation of RR can be amplified by the underrepresentation of credits with extended workout periods that start before or end after the observation window. This issue can be concrete when analyzing NPCs from the original creditor's perspective, where the recovery duration is not fixed and the focus is on the ultimate recovery. However, this issue is less relevant in our sample, as third-party collector data feature a contract-specific fixed recovery duration, the recovery efforts begin upon acquisition of the NPC, and the RR is measured at the end of this fixed period. Third-party collector data also helps maintain consistency by having the same collection agency oversee recovery processes across different categories of defaulted consumer credits, reducing potential biases associated with internal lender-specific procedures. Nevertheless, limitations remain, such as the absence of detailed workout cost data—shaped by the contractual terms between the lender and the collector—and the focus on the contract-specific recovery period. Therefore, similar to the case of credit cards in Bellotti and Crook (2012) and telecommunication-related credits handled by third-party collectors of Nazemi et al. (2022), the RR represents the outcome of the recovery procedure within a specified collection period relative to the Total Amount to Recover (TtR). Consequently, the RR definition for NPC$_i$ considered in this paper is

$$\mathrm{RR}_i = \frac{\text{sum of repayments}_i \text{ over the period } [\tau_i, \tau_i + \Delta_i]}{\mathrm{TtR}_i \text{ at } [\tau_i]}\,, \tag{1}$$

where the collection period starts in τ_i when the debt collector receives the recovery mandate and its length Δ_i can vary across credits, but cannot exceed one year. The TtR includes the principal, interests, recovery fees, and administrative costs imposed by the original lender for the defaulted credit. The gross recovery rate RR reflects the retrieval of the principal, potential interests, and additional fees established by the lender, but excludes third-party collector handling fees. These fees, typically comprising a fixed per-piece fee and a percentage commission on the recovered amount, are specified in the bilateral contract between the original lender and the third-party collector. It is the lender's responsibility to pay these fees to the third-party collector at the end of the collection period.[6]

The data covers the period from January 2007 to December 2019. We have access to the debtor's postal code and the date when the third-party agency was authorized to recover the defaulted credit to match each NPC to a specific region and province at a specific moment in time. The database offers a representative picture of the situation in Italy since all 20 regions and 101 out of 107 provinces are represented. Table 1 reports the descriptive statistics of the RRs across types of debtors and sectors, 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 defaulted credits originate from utility (35.04%) and telecommunication sectors (48.03%), financing (car loans, credit cards, leasing, 5.45%), banking (overdrafts and personal credits, 3.72%) and commercial credit (sales credit, 7.78%). NPCs originating from the telecommunication industry (Telco) display the lowest average RR of 19%, followed by the utility and commercial sectors, yielding 27% and 31%, respectively. NPCs originating from the banking and financing industry display higher average RRs (respectively, 48% and 47%).

Figure 1 confirms the usual bi-modal distribution of recovery rates previously observed in this field, with no-recovery ($RR = 0\%$) and full recovery ($RR = 100\%$) being the two most likely outcomes of the collection process (see, e.g., Gambetti et al., 2019; Nazemi et al., 2022). The Bernoulli-shaped distributions emphasize the high uncertainty inherent to consumer credit recovery rates. The only exception is the banking and financing sectors, which exhibit a third probability mass peak at 50%. The last section of Table 1 confirms previously documented evidence about the negative relationship between RRs and TtR (Nazemi et al., 2022). We identify four categories of debtors: male and female individuals, professionals, and unspecified and we also observe the age of individual debtors. We refer to Supplementary Material B for a geographical visualization of NPC-specific features.

3.2 Macroeconomic and social predictors

We use the debtor's postal code and the start date of the collection period to match each NPC to a specific region and province at a specific moment in time. To better explain the heterogeneity of the data across regions and highlight possible trends, we divide the regions into three macro-territories: North, Center, and South.[7] We have a total of 13 such variables that we can classify into four broad classes: business cycle, credit cycle, economic uncertainty, and social environment.

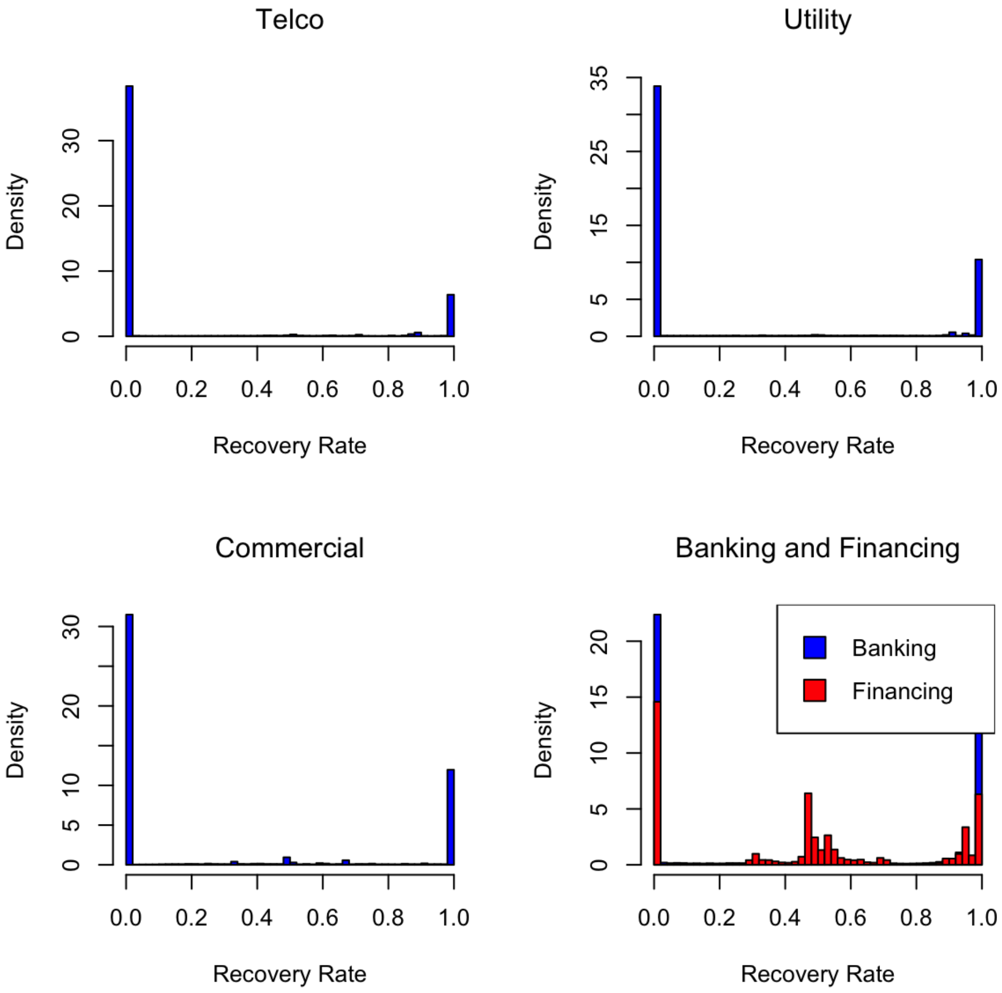
We begin with the four business cycle indicators. We consider the monthly seasonally adjusted unemployment rate and year-on-year growth rate of the Harmonized Index of Consumer Prices (HICP) as a proxy for inflation. Because of inconsistencies in the reporting of inflation and unemployment series at the regional or provincial level, we stick with these measures at the national level. Observe, however, that GDP is not available at the regional level. Therefore, we use the regional real gross value added (GVA) as a proxy for the regional GDP. Figure 2 illustrates the average real gross value added and yearly growth rate, highlighting significant economic differences between regions in terms of growth and log levels. Then we move to the four variables used to proxy the dynamics of the credit cycle. Public debt

---

[6] This definition eases comparison with previous works as a similar definition has been used in academic papers that we consider as benchmarks (See for example Bellotti and Crook, 2012; Nazemi et al., 2022).

[7] 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. More information available at: `https://www.istat.it/en/archivio/227202. 8`

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



and cost-of-debt to GDP ratios are used to evaluate credit conditions at both the national and regional levels. In Figure 3, the top panel shows the variations in the cost-of-debt to GDP (on the left-hand side) and debt-to-GDP ratios (on the right-hand side) across regions. Generally, the Center and the South have a higher level of debt relative to their GDP and also experience a higher cost of debt compared to the North. We finally incorporate the spread between Italian and German 10-year constant maturity sovereign bond yields. To capture the effect of economic uncertainty we include the uncertainty index of Baker et al. (2016) for Italy.

Table 1: Recovery rate descriptive statistics of NPCs by debtors type, Maximum Recovery Duration (maxRD) ranging from <30 days to 365 days, 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
maxRD	# obs	μ	σ	.10	.25	.50	.75	.90
<30	477,589	0.5	0.4	0	0	0.486	0.956	1
30-60	929,714	0.45	0.47	0	0	0.204	1	1
60-90	1,108,511	0.35	0.46	0	0	0	0.268	1
90-120	1,336,891	0.17	0.36	0	0	0	0	1
120-240	2,463,494	0.20	0.37	0	0	0	.0123	1
240-365	177,595	0.11	0.28	0	0	0	0	.600
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

Figure 2: Heterogeneity of Cost-of-debt and Debt to GDP ratios (top) and log-level of GDP and growth rate (bottom) at the region level. The regional GVA is used as a proxy for the GDP of the region.

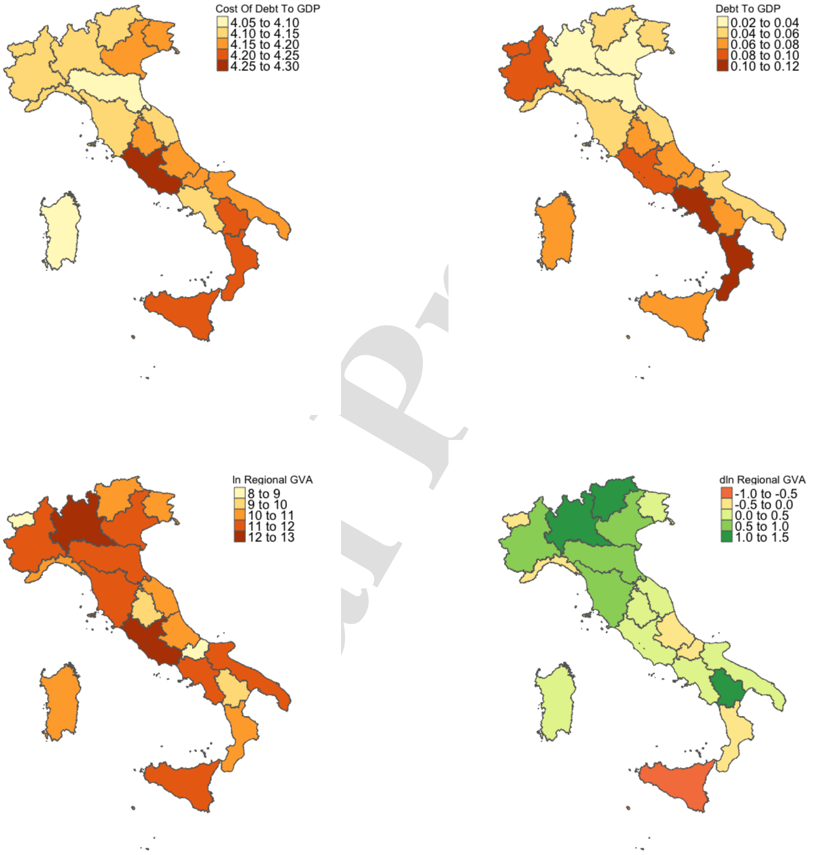
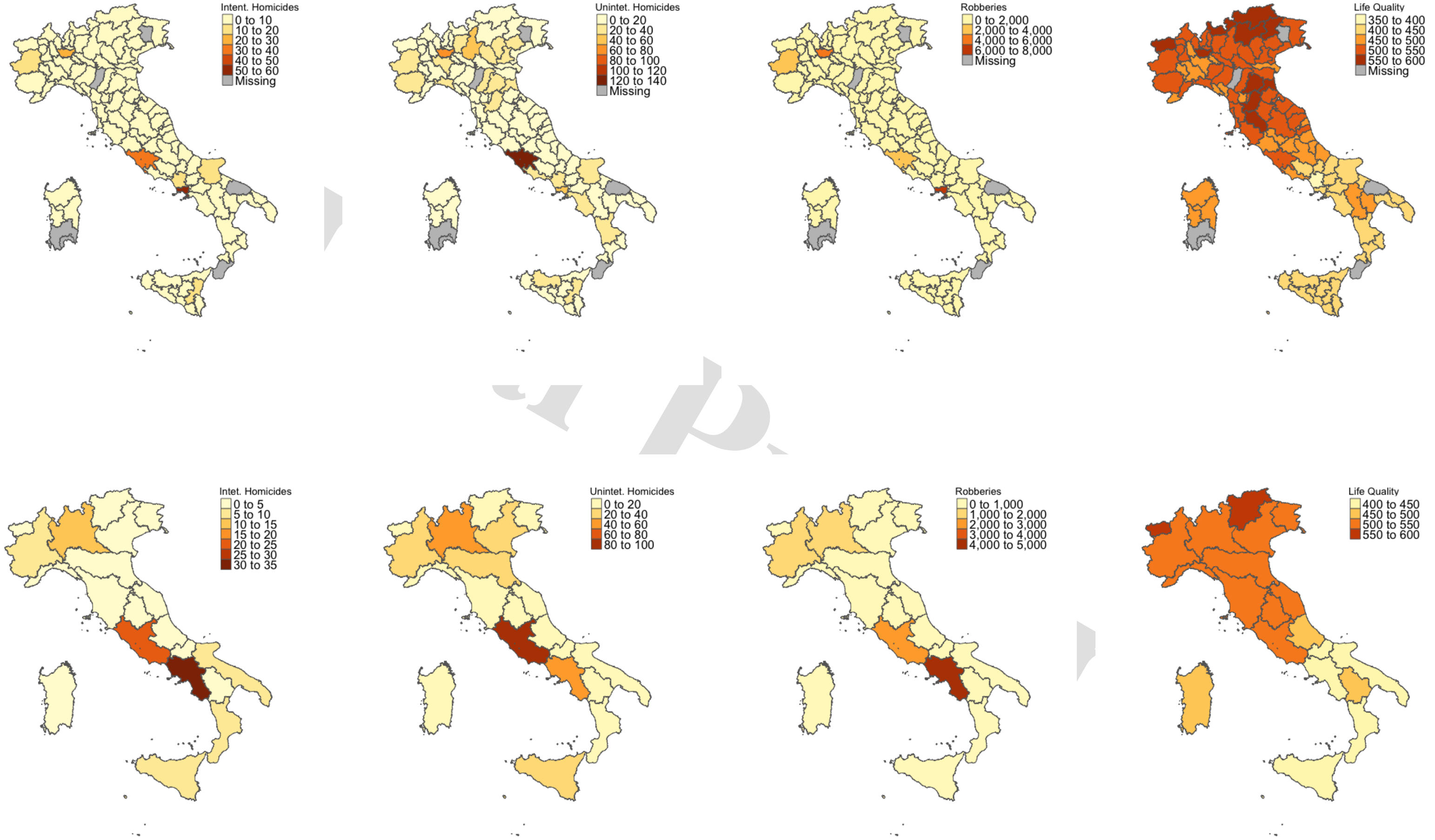


Figure 3: Heterogeneity of the number of intentional (left) and unintentional (center-left) homicides, robberies (center-right) and Life Quality Index (right) at province (top) and region (bottom) level.



The social environment can potentially affect the outcome of the credit collection process. For instance, higher levels of criminal activity, lower income, or less developed areas may directly affect the realized RR. To capture the socio-economic heterogeneity at the province level, we examine four variables: 1) the number of robberies, 2) intentional homicides, 3) unintentional homicides (all scaled per 100,000 inhabitants), and 4) the Life Quality Index published by the Italian financial newspaper *Il Sole 24 Ore*.[8] The top panels of Figure 3 reveal that Milan, Rome and Naples are the provinces with the highest numbers of robberies and intentional and unintentional homicides per 100,000 inhabitants. The right panels of Figure 3 display instead the variation in the Life Quality Index across regions and provinces of Italy. The data highlights that regions and provinces in the North and the Center tend to have higher average values of the Life Quality Index (538 and 521, respectively), compared to the South (419.95). 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. Table 2 indicates the sources for the MS indicators and their descriptive statistics, and also reports the descriptive statistics of contract-specific variables.

3.3 Portfolios of defaulted consumer credits

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

We consider two scenarios that mimic the day-to-day operations of debt collectors. In the first scenario (scenario 1), third-party collectors have access to the precise information of each individual loan in the pool, allowing them to create forecasting models using constituent-level data. In this case, they can forecast the recovery rates of individual NPCs forming the pool, and therefore the forecasting performance can be evaluated both at individual credit and portfolio levels. In the second scenario (scenario 2), we assume that third-party collectors can access NPC information communicated in some aggregate way, at the portfolio level.[9] Predictive models are thus trained based on a "portfolio-level predictor set", and their output directly provides the forecast of the recovery rate for the entire pool. In this scenario, forecasting performance can be evaluated only at the portfolio level.

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

---

[8] 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. For more information visit `https://lab24.ilsole24ore.com/qualita-della-vita/`

[9] This is standard practice in this kind of industry, as well as in the sector of Asset-Backed Securities.

Table 2: Descriptive statistics for macroeconomic and social (MS) variables, and contract-specific variables. The first part of the table lists macroeconomic and social variables, and the second part lists loan-specific variables. For MS indicators, the column Freq. indicates whether the frequency is monthly (M) or yearly (Y), while the column NUTS (Nomenclature of territorial units for statistics) indicates whether the variable is at the national (1), regional (2), or district (3) level. ISTAT stands for the Italian National Institute of Statistics (`https://www.istat.it/en/`) and BLB stands for Bloomberg.

Macroeconomic and Social Variables

Variable	Abrev.	Source	Mean	Std.Dev.	Min	Pctl. 25	Pctl. 75	Max	Freq.	NUTS
Regional GVA Mln	ln GVA	ISTAT	119414.5	96610.8	4103.1	64286.2	141279.0	344629.2	Y	2
Regional GVA dln	dln GVA	ISTAT	0.1	2.0	-8.6	-0.3	1.1	8.9	Y	2
National Unemployment	Un	ISTAT	11.8	5.9	2.7	7.3	18.4	23.4	M	1
EPU Italy	EPU	Baker et al. (2016)	111.0	33.0	31.7	89.5	130.0	241.0	M	1
HICP YoY	HCIP	ISTAT	0.9	0.8	-0.5	0.2	1.4	4.2	M	1
Regional Debt to GDP	R-DtGDP	ISTAT	0.1	0.0	0.0	0.0	0.1	0.1	Y	2
National Debt to GDP	N-DtGDP	ISTAT	129.8	10.0	103.9	132.5	134.8	135.4	Y	1
National Cost of Debt to GDP	N-CtGDP	ISTAT	4.1	0.5	3.6	3.8	4.6	5.2	Y	1
Spread ITA-GER 10Y bonds	Spread	BLB	171.5	65.7	21.1	130.7	205.3	460.3	M	1
Life Quality Index	LQ	ilsole24ore.com	490.2	66.6	360.0	425.0	545.6	641.5	Y	3
Intentional Homicides	INHom	ISTAT	11.0	16.5	0	1	13	109	Y	3
Unintentional Homicides	UNHom	ISTAT	30.9	31.8	0	10	40	138	Y	3
Robberies	Rob	ISTAT	1337.3	2202.4	4	91	1799	14045	Y	3

Contract-Specific Variables

Variable	Abrev.	Mean	Std.Dev.	Min	Pctl. 25	Pctl. 75	Max
Age	Age	55.335	17.572	21	43	67	100
Max Recovery Duration	maxRD	90.877	76.28	15	30	120	365
Gender: F	F	0.272	0.445	0			1
Gender: M	M	0.425	0.494	0			1
Gender: Professionals	Prof.	0.006	0.076	0			1
Recovery Rate	RR	0.344	0.43	0	0	0.91	1
Credit to Recover	TtR	727.48	1104.19	50.01	185.13	742.672	10,000
Principal to Total to Recover	PTtR	0.881	0.186	0	0.874	1	1
Banking	Bank.	0.2	0.4	0			1
Commercial	Comm	0.2	0.4	0			1
Financing	Fin.	0.2	0.4	0			1
Telecommunications	Telco	0.2	0.4	0			1
Utilities	Util.	0.2	0.4	0			1
Region: Abruzzo	ABR	0.026	0.158	0			1
Region: Basilicata	BAS	0.008	0.091	0			1
Region: Calabria	CAL	0.034	0.181	0			1
Region: Campania	CAM	0.145	0.352	0			1
Region: Emilia Romagna	EMR	0.055	0.227	0			1
Region: Friuli Venezia Giulia	FVG	0.012	0.111	0			1
Region: Liguria	LIG	0.027	0.161	0			1
Region: Lombardia	LOM	0.135	0.342	0			1
Region: Marche	MAR	0.017	0.131	0			1
Region: Molise	MOL	0.006	0.076	0			1
Region: Piemonte	PIE	0.067	0.25	0			1
Region: Puglia	PIG	0.045	0.207	0			1
Region: Sardegna	SAR	0.011	0.102	0			1
Region: Sicily	SIC	0.107	0.309	0			1
Region: Trentino Alto Adige	TAA	0.006	0.076	0			1
Region: Tuscany	TUS	0.055	0.228	0			1

We consider portfolios of 100 credits because this offers an acceptable compromise between keeping enough observations for the training (and testing) set and maintaining the portfolio large enough to reduce the relevance of individual credits at the portfolio level. Figure 10 in the Supplementary Material A displays the number of portfolios over time.

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

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

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

4 Methodology

When third-party collectors have precise information on the individual loans in the pool, they can leverage constituent-level data to estimate the forecasting models. Following the methodology of previous studies in the recovery rate forecasting literature (Nazemi et al., 2017, 2018; Nazemi and Fabozzi, 2018; Gambetti et al., 2022), we use stratified sampling to create training and testing sets with the same proportions of NPCs originating from different industries (telecommunication, utilities, banking, commercial loans, or consumer finance). To keep the computational cost of the study under control and, more importantly, to save enough observations in the test set for evaluating the forecasting performance at the portfolio level when using models trained using constituent-level data, we consider a 10-90% split between training and testing sets. This processing provides us with 124,193 individual NPCs for training and 1,080,562 individual NPCs for testing equally spread across industries. When evaluating forecasting performance at the portfolio level, we aggregate these individual NPCs in the test set into 10,800 portfolios, as detailed in Section 3.3.[10]

When third-party collectors can access information at the portfolio level only, we estimate the forecasting models and evaluate the performance using portfolio-level data only. In this case, we use the 6,493,794 observations to build 62,082 portfolios as indicated in Section 3.3 and consider a standard 70% - 30% split between training and testing sets. As a robustness check, we design a prediction exercise in which the third-party collector ensures that the training set only features portfolios preceding the ones included in the testing set. This approach (also known as "out-of-time") controls for potential structural breaks in the data-generating process which could potentially lead to a look-ahead bias in the predictions (Kalotay and Altman, 2016; Nazemi et al., 2022). In this exercise, the third-party collector lacks data on the specific loans forming the portfolios and uses portfolios of NPCs acquired before January 2017 for training, while credits acquired after January 2017 are left for testing. This corresponds to an approximately 50% - 50% split between training and testing sets permitting us to have enough observations to train data-intensive predictive models while keeping enough variability of yearly MS series.[11]

5 Forecasting methods and evaluation criteria

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

---

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

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

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

Table 3 lists the predictive algorithms and forecast combinations used in this study. For a detailed description of these algorithms, see Supplementary Material C, while for additional discussion on similar ML approaches in recovery rate forecasting, refer to Bellotti et al. (2021) and Gambetti et al. (2022). We will now provide a brief overview of the three forecast combination methods.

Table 3 reports the predictive algorithms and the forecast combinations employed in this study. We refer to the Supplementary Material C for a detailed description of the predictive algorithms employed, while the interested reader may refer to Bellotti et al. (2021) and Gambetti et al. (2022) for further discussion on the similar ML approaches used in recovery rate forecasting. We now proceed to describe the three forecast combinations (FC) methods briefly.

Table 3: List of considered algorithms and corresponding `R` packages.

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

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

We consider the equally weighted (i.e., simple average) forecast combination (EW FC), and two optimization-based model mixtures that are estimated from the data: a) the minimum-variance linear combination of forecast with nonnegative weights (Opt+ FC) and, b) the neural-network based nonlinear combination of forecasts (ANN FC). Assuming that the individual forecasts are unbiased[12], Opt+ FC minimizes the variance of the aggregate forecast error. In this case, restricting the weights to sum to one

---

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

keeps the aggregated forecast unbiased as well (Timmermann, 2006). Following Granger and Ramanathan (1984), with the sum-to-one and non-negativity constraints on the combining weights, the optimization problem becomes

$$\boldsymbol{w}^* := \underset{\boldsymbol{w}\,\in\,\mathcal{W}}{\operatorname{argmin}}\, \boldsymbol{w}^T \widehat{\boldsymbol{\Sigma}} \boldsymbol{w} \quad \text{where} \quad \mathcal{W} := \{\boldsymbol{w} \in \mathbb{R}^7 \mid \mathbf{1}^T \boldsymbol{w} \;=\; 1, \min \boldsymbol{w} \geq 0\}\,, \tag{2}$$

where $\widehat{\boldsymbol{\Sigma}}$ is the $7 \times 7-$sample covariance matrix of the models' prediction errors. This problem can be solved efficiently using quadratic programming or imposing the Karush-Kuhn-Tucker conditions. Restricting the weights to be non-negative and to sum to one implicitly selects which forecasts to combine and reduces the estimation error by inducing a shrinkage-like effect on the covariance matrix of forecast errors (Jagannathan and Ma, 2003).[13]

With ANN FC, we explore the relevance of nonlinear weighting schemes for the forecast combination. Specifically, we opt for averaged neural networks that take the forecasts produced by RF, GBRT, XGBt, MARS, B-MARS, Beta, and OLS linear regression models as inputs and aggregate these into one prediction. ANN FC is an ensemble of one-hidden-layer feed-forward neural networks (Ripley, 1996) and with the *sigmoid* activation function. The ensemble counts 20 neural networks that are initialized by using different starting values for the parameters to estimate and that include a *decay* factor to penalize large coefficients.[14] These features moderate the risk of over-fitting.

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

As performance criteria, we consider the most popular measures for forecasting models, namely, the RMSE, the MAE, and the R^2, defined as

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

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. A negative R^2 signals that the forecast method at hand under-performs the strategy considering the average RR in the training set as forecast. We compute the averages of out-of-sample performance measures for the models via bootstrap, considering samples of 100 observations and repeating the procedure 10,000 times. As in Nazemi et al. (2022), we use the Mann–Whitney test to assess whether the differences in performance when including MS indicators predictors are statistically significant. Finally, we also consider the model confidence set (hereafter MCS, Hansen et al., 2011), which

---

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

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

tests whether a subset of methods enters jointly in the superior set of models by repeatedly testing the null hypothesis of equal predictive performance for a given significance level α.[15]

6 Results

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

6.1 Forecasting performance

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

We first analyse the difference in performance when looking at the individual *vs.* portfolio level. Comparing panels a) and b), we observe a substantial enhancement in predictive accuracy when working with portfolios. For instance, if we choose Opt+ FC when MS predictors are excluded, the R^2 value rises from 28.27% for individual loans to 51.66 % in the context of portfolios. This suggests that the idiosyncratic risk of individual credits' recovery rates tends to be diversified away. This is illustrated in Figure 4, which shows how the aggregation into portfolios impact the RR distribution: it becomes uni-modal and its variance shrinks.

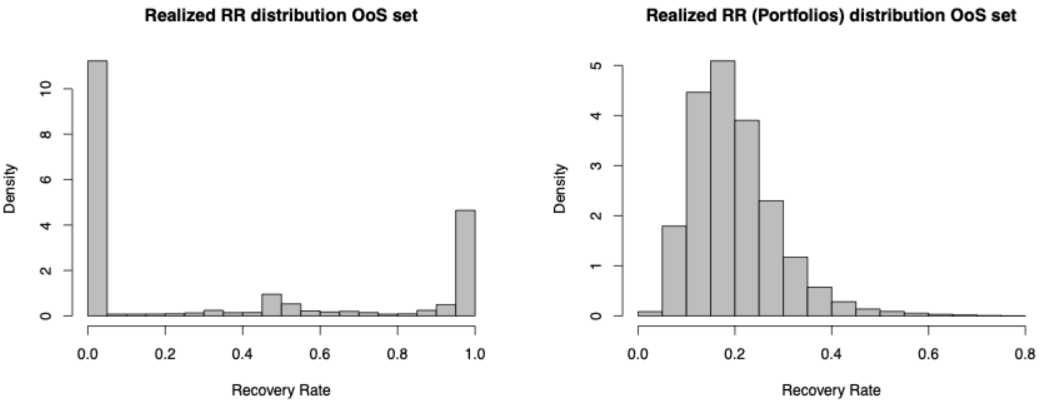
This is consistent with the theoretical result on the asymptotic distribution of the averaged RR of an equally weighted infinitely granular pool, where defaults are controlled by a Gaussian copula (see the discussion in the Supplementary Material E). Comparing panels b) and c), we study the benefit of possessing a constituents-level information set when forecasting aggregate RRs. The forecasting performance significantly deteriorates in absence of detailed information on the loans in the portfolio. For instance, if we again choose Opt+ FC when MS predictors are excluded, the R^2 value decreases from 51.66% to 32.40% when one only possesses portfolio-level information. In addition, Table 4 suggests that forecast combination techniques provide the most effective approach for predicting RRs. In particular, the use of artificial neural networks as a nonlinear combination strategy (ANN FC) is encouraging since it is consistently included in the superior set of models and attains the lowest MAE and RMSE values.

Let us now study the impact of MS indicators on the performance. The last three columns of Table 4 display the percentage differences in performance measures when MS indicators are excluded from the predictor set. Performance significantly deteriorates: this is evident when forecasting the recovery rates of defaulted individual credits (panel a), but it is even more apparent at the portfolio level (panels b and c).

---

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

Figure 4: Distributions of the realized recovery rate for defaulted individual credits (left) and portfolios of consumer credits (right).



When examining individual loans, the maximum improvement in R^2 is approximately 3 percentage points (XGBt in panel a) which is small but not negligible in this literature. The improvement in portfolios is even more pronounced. For example, including MS predictors in OLS (panel b) increases the R^2 by more than 10 percentage points and by more than 6 percentage points in the case of RF (panel c). Finally, we control for potential look-ahead bias in the models adopting an out-of-time setup. For conciseness, we report the performance associated with the portfolio level in scenario 2. Panel d) reports the performance of forecasting models trained exclusively on portfolio-level information and NPCs obtained before January 2017. Compared to the random training and testing setup in panel c), we note that performance generally worsens. For instance, the R^2 value drops from approximately 37% to 28% for the ANN FC. Yet, the positive impact of incorporating MS indicators remains clear. Specifically, nine out of ten models improve their RMSE and R^2, and eight out of ten improve the MAE. The size of this improvement is particularly noteworthy. For instance, the ANN FC model sees a rise of over 7 percentage points in its R^2, while the Beta and OLS exhibit an increase of more than 15 p.p.

Table 4: Performance measures for each specification; best scores are shown in bold. The last 3 columns indicate the loss differentials when excluding MS and the respective significance levels for the Mann-Whitney test with ∗∗∗, ∗∗, and ∗, for the 1%, 5%, and 10% levels. We report the superior set of models at the 5% significance level and the L2 loss denoting whether the model include the MS or not with the acronyms *MS* or *no MS*.

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

6.2 Variable importance and local explanatory power

Variable importance (VI) plots indicate which variables are most useful for predicting the response variable. Within the RF forecasting framework, the RMSE is computed on the out-of-bag data for each tree and then recomputed after permuting a predictive variable. The differences are averaged and scaled to have a maximum value of 100 (Kuhn, 2007). For gradient-boosted trees, this method uses the same approach as a single tree but sums up the importance metrics over each boosting iteration.

We report the VI plots for RF (left panel) and GBRT (right panel) in Figure 5. We can highlight two important aspects. First, the VI plots for both GBRT and RF highlight the significance of sector (Banking in particular), TtR, maxRD, and the proportion of NPCs with debtors living in specific regions as among the most relevant contract-specific variables. Second, among the 43 predictors, N-CtGDP, N-DtGDP, the spread between Italian and German 10Y Sovereign bond yields (Spread), the Economic Policy Uncertainty Index (EPU), the inflation (HICP), and the Life Quality Index (LQ), are in the top 20 most relevant features within the RF predictive framework. In the GBRT model, unemployment (Un), N-CtGDP, N-DtGDP, the EPU, and Spread are in the top 10. More importantly, macroeconomic variables such as EPU, N-CtGDP, and Spread consistently rank higher than contract-specific variables like Age, albeit the importance of the latter has been highlighted in previous studies (see for example Table 5 in Nazemi et al., 2022).

To complete the picture, we provide in Figure 6 the mean of the absolute value of the SHAP (SHapley Additive exPlanations) values within the GBRT and XGBt framework corresponding to the predictive variables[16]. Similarly to the VI plot provided in Figure 5, the larger the contribution, the more relevant the predictor is in driving the forecasts in the test set. Both figures identify that MS indicators like EPU, N-CtGDP, and N-DtGDP are important drivers of the RR forecasts. Interestingly, they are ranked before Age or Gender (M/F), which are two contract-specific variables that have been previously identified as important determinants of RR. Overall, this explanatory analysis reinforces our previous findings on the importance of MS variables in improving the performance of models forecasting RRs for consumer credits. While contract-specific factors remain the primary drivers, some MS indicators demonstrate higher explanatory power than certain idiosyncratic variables identified in earlier studies. This supports our conclusion that MS variables, though not substitutes for contract-specific features, provide complementary information that helps explain variations in RR across different economic environments.

6.3 Visualizing the Accumulated Local Effects and Lime Plots

To examine the average effect of each predictor, we use ALE plots which rely on a localization to analyse model predictions.[17] The vertical axes measure the differences for the average prediction, while the horizontal axes the domain of the corresponding variable.

---

[16] The Shapley values are computed following Lundberg et al. (2020), focusing on the GBRT optimized with stochastic gradient boosting and XGBt to limit computational cost.

[17] Another global effects visualization technique is Partial Dependence (PD) plots, but they tend to be more computationally expensive and may produce inaccurate results when predictors are highly correlated. For a more in-depth discussion of ALE, PD, and marginal plots, we refer readers to Apley and Zhu (2020). Note that the models we consider assume different functional forms for connecting RR to its determinants (e.g., linear models vs. trees) or use different training algorithms (e.g., random forest vs. gradient boosting), leading to quantitatively different relationship shapes. However, our discussion highlights that the ALE plots may display qualitatively similar patterns. As a robustness check, we also perform an analysis of local effects using Local Interpretable Agnostic Model Effect methods. The results are in line with our previous findings and we refer to the Supplementary Material for the corresponding discussion.

Figure 5: Variable Importance plots: relevance of predictive variables.

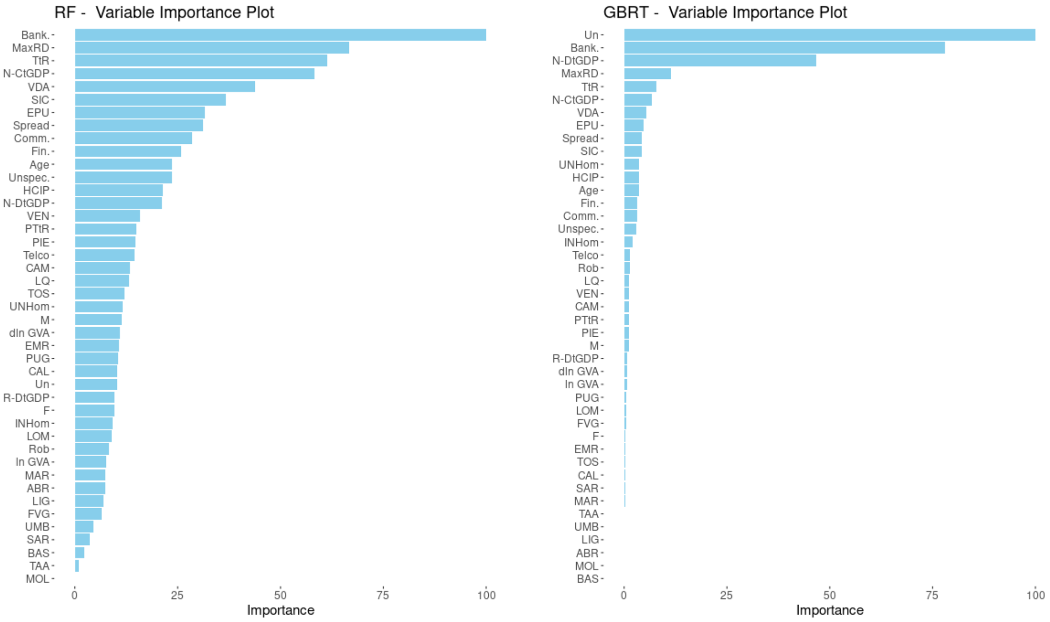
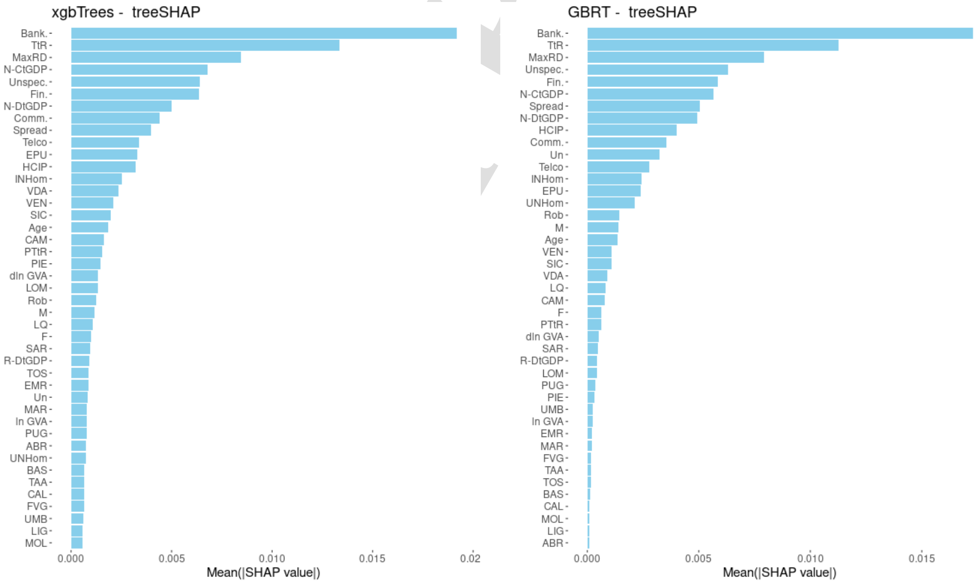


Figure 6: The average of the absolute value of SHAP (SHapley Additive exPlanations) values across observations in the testing set for the GBRT and XGBt algorithms.



We analyse whether the industry originating the credit affects average recovery rates. ALE plots in Figure 7 confirm what we have already observed in Table 1. Setting the utility sector as a baseline, we see that models predict greater recovery rates when the proportion of credits from the banking sector increases (see Banking). We observe a similar but less accentuated effect related to the quota of credits from the financing industry (see Financing) and commercial credits (see Commercial); conversely, the greater the proportion of defaulted telecommunication credits in the portfolio, the lower the expected recovery rate. This highlights that findings about NPC RRs related to a specific sector cannot be easily generalized to other industries, in contrast with what previous research suggested (Nazemi et al., 2022).

For example, the greater expected recovery rate associated with the Banking quota may reflect a better screening ability or be consistent with a different debtor behavior (e.g., the fear of being refused future loans from banks).

Figure 7 also illustrates the impact of some loan-specific variables. A greater proportion of interest and ancillary fees in the total amount to recover leads to a lower expected recovery rate (see Principal to Total to Recover). We confirm the negative correlation between the total amount to recover and the recovery rate (see Credit to Recover) previously observed in other papers (Bellotti et al., 2021; Nazemi et al., 2022). Interestingly, being granted more time to perform the collection does not seem to enhance expected recovery rates, on the contrary (see max Recovery Duration).

Figure 8 focuses on macroeconomic indicators. We notice that higher levels of public debt interest are associated with lower expected recovery rates (see National Cost of Debt to GDP). In addition, low uncertainty seems to be associated with higher expected RRs (see EPU in Figure 8). Among the MS variables, the EPU has the largest effect, extending the findings of Gambetti et al. (2019) to defaulted consumer credits. National inflation seems positively associated with higher expected RRs, and the estimated ALE varies across models (see HICP YoY). However, it is worth noting that within our sample, the inflation level never went above 4% and we cannot properly assess the effect of high inflation rates. Finally, a high level of real regional output is also associated with a greater expected recovery rate. The relationship is highly nonlinear, and the size of the effect is modest and only captured by gradient-boosted trees (see Regional GVA).

While global methods like ALE plots examine the relationship between the target variable and the feature across the entire sample space, local interpretations help us understand model predictions for a single row of data or a group of similar rows. Therefore, we also include LIME (Local Interpretable Model-agnostic Explanations) plots to complement. LIME operates under the assumption that every complex model is linear on a local scale, allowing us to fit a simple (linear) model around a single observation that will mimic the global model's behavior in that locality.

Figure 9 displays the results for three representative classes of models: RF, GBRT and B-MARS. These plots distinguish the contributions of each feature into positive and negative effects on the predicted RR. For example, the top left and bottom panels show the positive and negative contributions of the predictive variables in explaining the recovery rate forecasts produced by the RF model. The main limitation of LIME plots is the assumption that the forecast can be *linearly* explained by the set of features. The low R^2 values associated with the linear regression model aiming to explain the RR forecasts (e.g., 0.302 for GBRT and 0.491 for RF) may indicate a violation of this assumption, limiting the reliability of LIME plots in identifying the local drivers of the RR forecasts. Nevertheless, we confirm that *Spread*, *ln GVA*, *Un*, and *N-DtGDP* significantly contribute (both positively and negatively) to explaining RR forecasts produced by RF and GBRT, supporting the evidence from the variable importance plots.

Figure 7: ALE plots: impact of the sector.

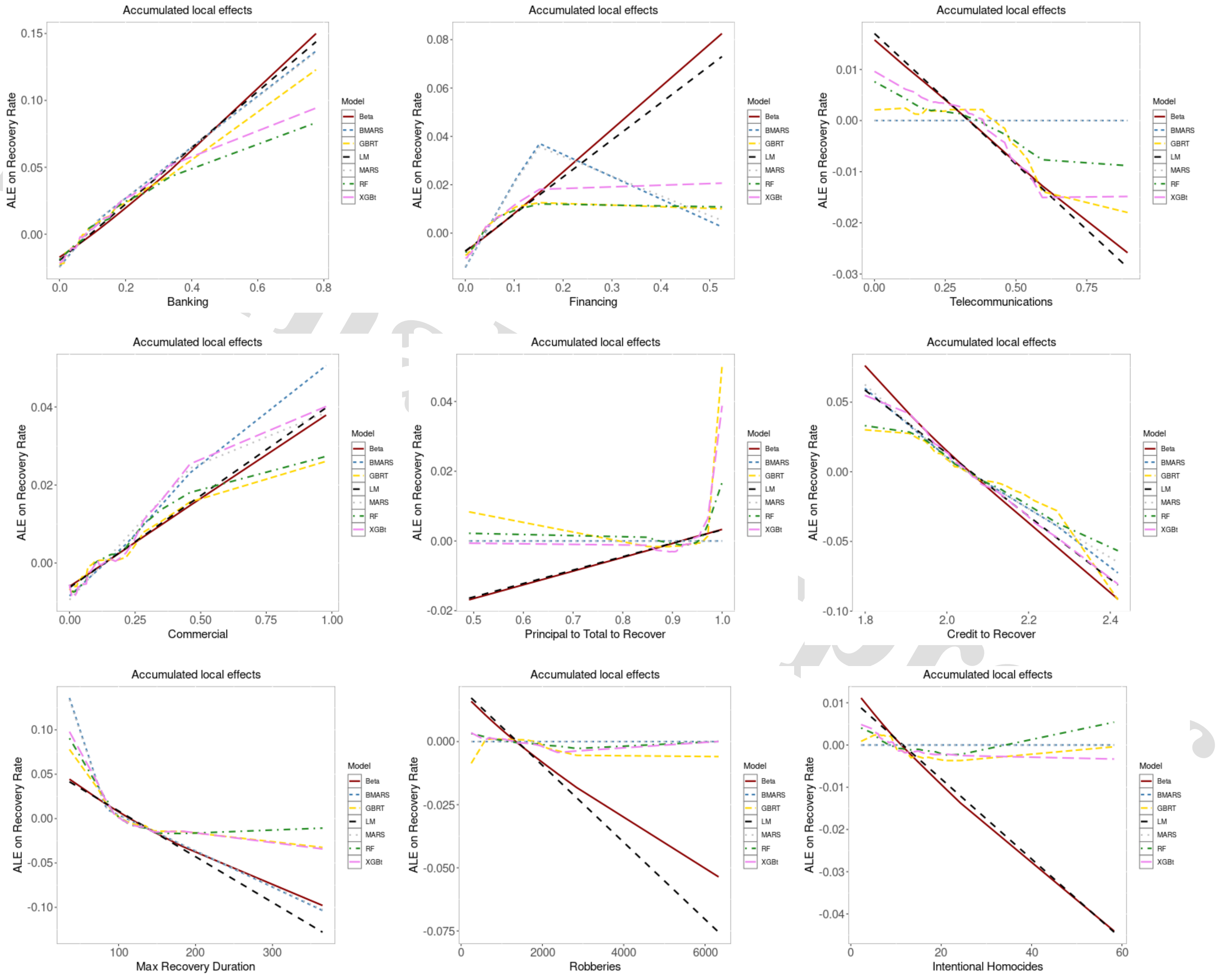


Figure 8: Accumulate Local Effect Plots.

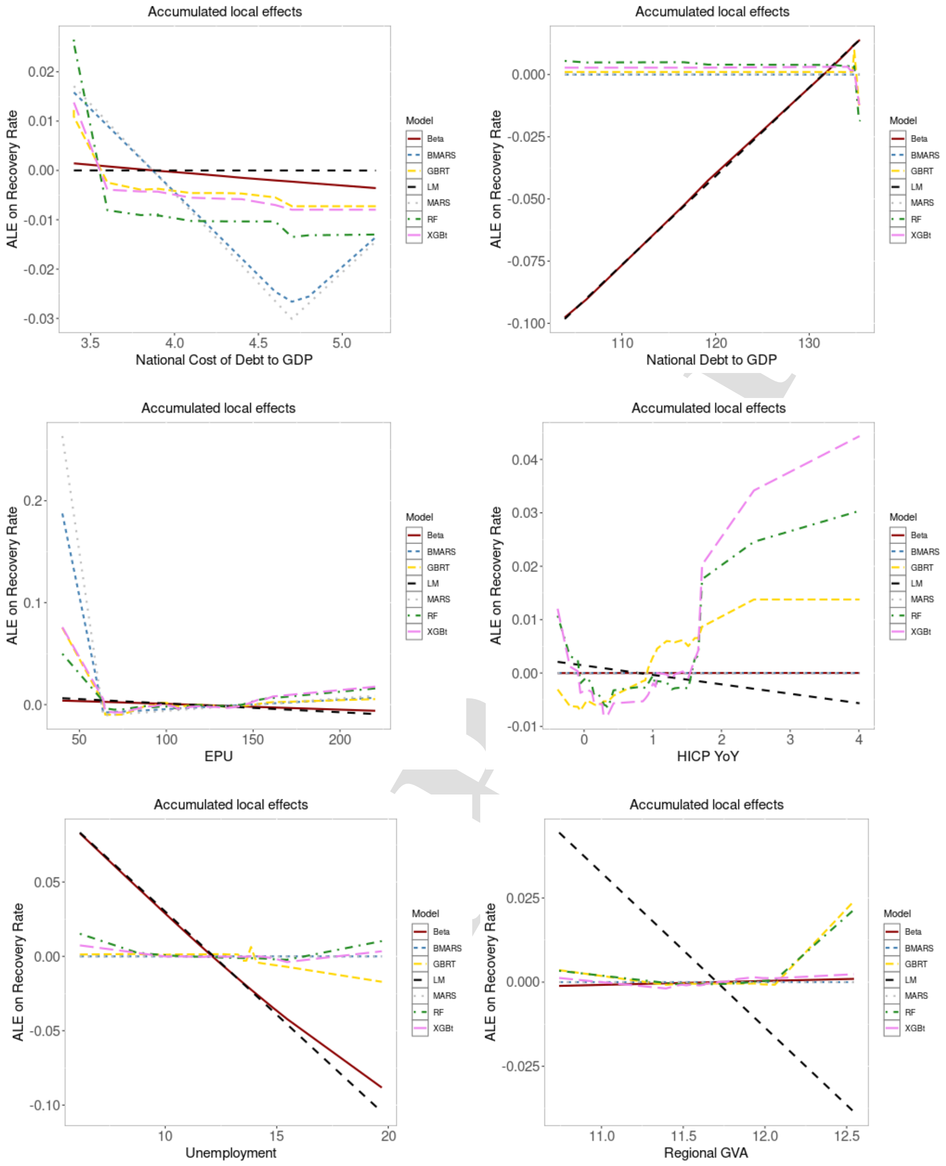
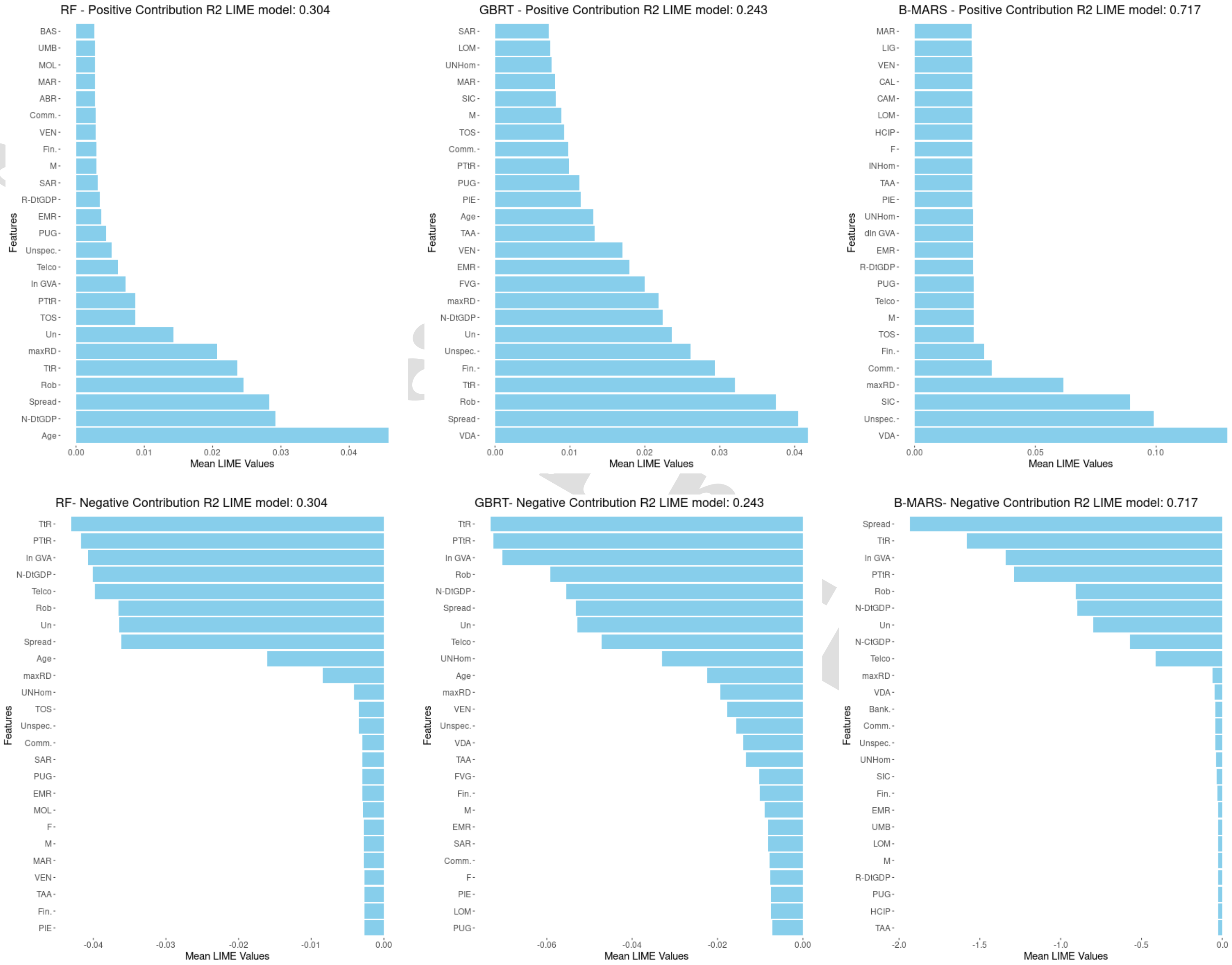


Figure 9: LIME plots across observations in the testing set distinguishing between positive and negative contributions.



7 Conclusions

We study the drivers of realized losses in defaulted consumer credits from the perspective of stakeholders involved in the recovery process. To make informed decisions—whether to initiate a collection procedure, determine third-party debt collection fees, or price an NPC for sale—every stakeholder needs to forecast the recovery rates. In contrast to previous studies, our findings demonstrate the relevance of including macroeconomic variables and social indicators for enhancing forecasting performance. Our results hold irrespective of the forecasting model or whether we aim at forecasting the recovery rate for an individual loan or a portfolio of loans, and are relatively stronger in the latter case. This result might be useful to industry participants as it is closer to their day-to-day business operations.

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

In light of recent studies highlighting the potential of explainable AI techniques, the analysis of accumulated local effects shows that the relationship between expected recovery rates and predictors related to the dynamics of the business cycle is intuitive: deteriorating credit conditions, indicated by lower real activity, a rising cost-to-GDP ratio, and increasing economic uncertainty, are associated with lower expected recovery rates. Locally using Shapley values, we find that contract-specific features like the maximum recovery duration and the sector of the defaulted credit remain crucial drivers; however, macroeconomic and social indicators provide important complementary information that explains variations in LGD across different economic conditions.

Our study focuses on the role of contract-specific features and aggregate macro information. However, it might be worth investigating other potential conditioning factors. One notable example is weather-related events. In 2019, extreme events led to economic losses equivalent to 1% of GDP in the euro area (ECB, 2021). Extreme climate and weather events, such as floods, wildfires, and hurricanes, negatively impact GDP by damaging property and physical capital, and by affecting investments and financial institutions. Additional hazards, like water and heat stress, diminish labor and agricultural productivity, disrupt logistics, and cause economic activities to relocate. These physical hazards have long-term effects on GDP and employment by causing sustained production losses and redirecting investment capital toward reconstruction efforts (Meier et al., 2023). Climate change is expected to increase the frequency and intensity of these hazards, thereby heightening physical risks. While incorporating macroeconomic and social indicators may only indirectly and partially capture the effects of natural disasters, the primary challenge is to identify the mechanisms through which climate risk factors impact the recovery process and to directly quantify the economic impact of natural disasters on recovery rates. We leave this important question for future research.

Acknowledgments

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

References

Altman, E., B. Brady, A. Resti, and A. Sironi (2005). The link between default and recovery rates: theory, empirical evidence, and implications. *The Journal of Business 78*(6), 2203–2227.

Apley, D. W. and J. Zhu (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society Series B 82*(4), 1059–1086.

Atiya, A. F. (2020). Why does forecast combination work so well? *International Journal of Forecasting 36*(1), 197–200.

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics 131*(4), 1593–1636.

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

Bastos, J. A. and S. M. Matos (2022). Explainable models of credit losses. *European Journal of Operational Research 301*(1), 386–394.

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

Bellotti, A., D. Brigo, P. Gambetti, and F. Vrins (2021). Forecasting recovery rates on non-performing loans with machine learning. *International Journal of Forecasting 37*(1), 428–444.

Bellotti, T. and J. Crook (2012). Loss given default models incorporating macroeconomic variables for credit cards. *International Journal of Forecasting 28*(1), 171–182.

Breiman, L. (1996). Bagging predictors. *Machine Learning 24*(2), 123–140.

Breiman, L. (2001). Random forests. *Machine Learning 45*(1), 5–32.

Bussmann, N., P. Giudici, D. Marinelli, and J. Papenbrock (2020). Explainable machine learning in credit risk management. *Computational Economics 57*(1), 203–216.

Calabrese, R. (2012). Predicting bank loan recovery rates with a mixed continuous-discrete model. *Applied Stochastic Models in Business and Industry 30*(2), 99–114.

Carvalho, P. V., J. D. Curto, and R. Primor (2020). Macroeconomic determinants of credit risk: Evidence from the eurozone. *International Journal of Finance & Economics 27*(2), 2054–2072.

Caselli, S., S. Gatti, and F. Querci (2008). The sensitivity of the loss given default rate to systematic risk: New empirical evidence on bank loans. *Journal of Financial Services Research 34*(1), 1–34.

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

Dermine, J. and C. N. de Carvalho (2006). Bank loan losses-given-default: A case study. *Journal of Banking & Finance 30*(4), 1219–1243.

Djeundje, V. B. and J. Crook (2018). Incorporating heterogeneity and macroeconomic variables into multi-state delinquency models for credit cards. *European Journal of Operational Research 271*(2), 697–709.

ECB (2021). Climate change and financial stability. *Financial Stability Review*, 58–61.

European Central Bank (2016). Bank lending survey for the euro area: glossary. Technical report.

Fedaseyeu, V. (2020). Debt collection agencies and the supply of consumer credit. *Journal of Financial Economics 138*(1), 193–221.

Ferrari, S. and F. Cribari-Neto (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics 31*(7), 799–815.

Foglia, M. (2022). Non-performing loans and macroeconomics factors: The italian case. *Risks 10*(1).

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics 19*(1), 1–67.

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

Gambetti, P., G. Gauthier, and F. 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). A risk-factor model foundation for ratings-based bank capital rules. *Journal of Financial Intermediation 12*(3), 199–232.

Granger, C. W. J. and R. Ramanathan (1984). Improved methods of combining forecasts. *Journal of Forecasting 3*(2), 197–204.

Gürtler, M. and M. Hibbeln (2013). Improvements in loss given default forecasts for bank loans. *Journal of Banking & Finance 37*(7), 2354–2366.

Hacht, S. and R. Zagst (2010). Loan recovery determinants – a pan-european study. *SSRN Electronic Journal*.

Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica 79*(2), 453–497.

Höcht, S., A. Min, J. Wieczorek, and R. Zagst (2022). Explaining aggregated recovery rates. *Risks 10*(1), 18.

Hurlin, C., J. Leymarie, and A. Patin (2018). Loss functions for loss given default model comparison. *European Journal of Operational Research 268*(1), 348–360.

Jagannathan, R. and T. Ma (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance 58*(4), 1651–1683.

Kalotay, E. A. and E. I. Altman (2016). Intertemporal forecasts of defaulted bond recoveries and portfolio losses. *Review of Finance 21*(1), 433–463.

Kellner, R., M. Nagl, and D. Rösch (2022). Opening the black box – quantile neural networks for loss given default prediction. *Journal of Banking & Finance 134*, 106334.

Konecny, T., J. Seidler, A. Belyaeva, and K. Belyaev (2017). The time dimension of the links between loss given default and the macroeconomy. *Czech Journal of Economics and Finance 67*(6), 462–491.

Kuhn, M. (2007). Variable importance using the caret package.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software, Articles 28*(5), 1–26.

Kuhn, M. and K. Johnson (2013). *Applied Predictive Modeling.* New York. Springer.

Leow, M., C. Mues, and L. Thomas (2014). The economy and loss given default: evidence from two UK retail lending data sets. *Journal of the Operational Research Society 65*(3), 363–375.

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

Lundberg, S. M., G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence 2*(1), 56–67.

Malik, M. and L. C. Thomas (2010). Modelling credit risk of portfolio of consumer loans. *Journal of the Operational Research Society 61*(3), 411–420.

Meier, S., R. J. Elliott, and E. Strobl (2023). The regional economic impact of wildfires: Evidence from southern europe. *Journal of Environmental Economics and Management 118*, 102787.

Mora, N. (2012). What determines creditor recovery rates? *Federal Reserve Bank of Kansas City in its Journal Economic Review 79*(2), 109–123.

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

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

Nazemi, A. and F. J. Fabozzi (2024). Interpretable machine learning for creditor recovery rates. *Journal of Banking & Finance 164*, 107187.

Nazemi, A., K. Heidenreich, and F. J. Fabozzi (2018). Improving corporate bond recovery rate prediction using multi-factor support vector regressions. *European Journal of Operational Research 271*(2), 664–675.

Nazemi, A., F. F. Pour, K. Heidenreich, and F. J. Fabozzi (2017). Fuzzy decision fusion approach for loss-given-default modeling. *European Journal of Operational Research 262*(2), 780–791.

Nazemi, A., H. Rezazadeh, F. J. Fabozzi, and M. Höchstötter (2022). Deep learning for modeling the collection rate for third-party buyers. *International Journal of Forecasting 38*(1), 240–252.

Probst, P. and A.-L. Boulesteix (2017). To tune or not to tune the number of trees in random forest. *Journal of Machine Learning Research 18*(1), 6673–6690.

Qi, M. and X. Zhao (2011). Comparison of modeling methods for loss given default. *Journal of Banking & Finance 35*(11), 2842–2855.

Rapisarda, G. and D. Echeverry (2013). A nonparametric approach to incorporating incomplete workouts into loss given default estimates. *The Journal of Credit Risk 9*(2), 47–61.

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.

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

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

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

Highlights

- We study drivers of losses in defaulted consumer credits.
- Adding macroeconomic and social variables improves recovery rate forecasting.
- Non-linear forecast combinations using ANNs offer the best forecasts.
- ALE and SHAP values reveal relationships between recovery rates and key predictors.
- Lower output, rising public debt costs, greater uncertainty are linked to higher LGD.

Business cycle and realized losses in the consumer credit industry

Walter Distaso, Francesco Roccazzella, Frédéric Vrins

PII: S0377-2217(24)00968-8
DOI: https://doi.org/10.1016/j.ejor.2024.12.026
Reference: EOR 19336

To appear in: *European Journal of Operational Research*

Received date : 23 August 2024
Accepted date : 16 December 2024

Please cite this article as: W. Distaso, F. Roccazzella and F. Vrins, Business cycle and realized losses in the consumer credit industry. *European Journal of Operational Research* (2024), doi: https://doi.org/10.1016/j.ejor.2024.12.026.