DBA
IN MANAGEMENT AND TECHNOLOGY

Universidad Pontificia Comillas

Doctorate of Business Administration in Management and Technology

DEBT SEGMENTATION FOR COLLECTIONS AND RECOVERIES MANAGEMENT IN THE SPANISH MARKET. IMPACT OF A CRISIS: COVID-19 CASE STUDY

Author: Antón Alfaya González
Director: Dr. Álvaro Caballo Trebol

MADRID | February 2026

COMILLAS UNIVERSIDAD PONTIFICIA ICADE

COMILLAS UNIVERSIDAD PONTIFICIA ICAI

To Carolina, Sabela, Breogán, and Anxo, for being constant sources of inspiration, motivation, and support.

To my parents, Antonio and Fernanda, with and because of them all started.

A Carolina, Sabela, Breogán y Anxo, por ser una fuente constante de inspiración, motivación y apoyo.

A mis padres, Antonio y Fernanda, con y por ellos empezó todo.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the contributions and support of many people and institutions, to whom I extend my deep and sincere gratitude.

First, to David Tornero Suárez, who helped me develop the statistical models, and without whom, this thesis would not have been possible. His generous support and enriching comments have been fundamental to the development of this thesis and have been of immeasurable value to me.

I extend my thanks to Dr. Álvaro Caballo Trebol, my thesis advisor, who, with great generosity, patience, and insight, has tirelessly guided and supported me throughout the research process, refining the methodology and illuminating the path. I will be eternally grateful for his support, advice, and guidance.

I especially want to thank the Investment Firm and its staff for providing me with access to valuable information, for their assistance, and for their recommendations throughout the research.

I must also thank the entire community and the staff of the Pontificia Comillas University and its DBA program. In particular, Dr. Carlos Bellón Núñez-Mera for his support and inspiration throughout the program, as well as for his valuable comments and suggestions on the research and thesis; and Dr. Teresa Corzo Santamaría for her beneficial observations on the developed thesis. And also my DBA colleagues for sharing and enriching this experience.

Finally, on a personal level, this thesis would not have been possible without the support and encouragement of my family and friends. To my parents, for the upbringing and education I received with love and care, grounded in values of solidarity, generosity, effort, empathy, and a continuous curiosity for learning. To my siblings and friends, for always being there and being an endless source of energy.

And, above all, to Carolina, Sabela, Breogán, and Anxo, for your patience, understanding, tolerance, support, care, and unconditional love throughout the entire process. You have been a constant source of motivation and inspiration throughout the completion of the thesis, and this achievement is yours as well.

AGRADECIMIENTOS

"Prometo estarte agradecido"
R. Mercado

Esta tesis no sería posible sin la contribución y el apoyo de muchas personas e instituciones, a quienes quiero expresar un profundo y sincero agradecimiento.

En primer lugar, a David Tornero Suárez, quien me ha ayudado a construir los modelos estadísticos, y sin el cual esta tesis no habría podido tener lugar. Su apoyo generoso y sus comentarios enriquecedores han sido fundamentales para el desarrollo de esta tesis y de un valor incalculable para mí.

Extiendo los agradecimientos al Dr. Álvaro Caballo Trebol, mi director de tesis, quien, con gran generosidad, paciencia y clarividencia, me ha guiado y apoyado incansablemente en el desarrollo de la investigación, afinando la metodología a emplear y aportando luz a lo largo de todo el proceso. Le estaré eternamente agradecido por su apoyo, asesoría y guía.

Quiero agradecer especialmente a la Firma de Inversión y a su personal por darme acceso a información de gran valor, por su ayuda y por sus valiosas recomendaciones durante toda la investigación.

Debo agradecer a toda la comunidad y al personal de la Universidad Pontificia Comillas y de su programa de DBA. En especial, al Dr. Carlos Bellón Núñez-Mera, por su apoyo e inspiración a lo largo del programa, así como por sus valiosos comentarios y sugerencias sobre la investigación y la tesis; y a la Dra. Teresa Corzo Santamaría, por sus provechosas observaciones sobre la tesis desarrollada. Y también a mis compañeras y compañeros de DBA por compartir y enriquecer esta experiencia.

Finalmente, en el ámbito personal, esta tesis no habría sido posible sin el apoyo y aliento de mis familiares y amigos. A mis padres, por la formación y educación recibidas con cariño y amor, y basadas en valores de solidaridad, generosidad, esfuerzo, empatía y una curiosidad continua por el aprendizaje. A mis hermanos y amigos, por estar siempre ahí y ser una fuente de energía inagotable.

Y, por encima de todo, a Carolina, Sabela, Breogán y Anxo, por vuestra paciencia, comprensión, tolerancia, apoyo, cariño y amor incondicionales durante todo el proceso. Habéis sido una motivación e inspiración constantes para llevar a cabo la tesis, y este logro es vuestro también.

ABSTRACT

Collections and recoveries (C&R) activities are essential for effective credit risk management, particularly in light of post-2007-2008 financial crisis regulations. Despite their importance, this area is often underexplored. Accurate unpaid debt segmentation is crucial for optimizing collection strategies and assessing collection potential. Our objective is to identify the key determinants influencing recovery rates for unsecured debt portfolios that banks and financial institutions have written off, and to assess how these determinants were affected by the COVID-19 crisis. Specifically, we aim to identify which variables significantly predict short-term repayment behaviour before, during, and after the pandemic. To this end, we developed statistical models utilizing the CHAID classification tree methodology to identify the most discriminating variables. Our findings indicate that the COVID-19 crisis had a notable impact, causing significant alterations in these variables. The insights from this research are expected to be valuable to key stakeholders in the C&R sector, including creditors, servicers, and regulators/supervisors. In this research, we utilized a database from a prominent Investment Firm, containing information on 900.000 accounts of unsecured debt portfolios sourced from the Spanish non-performing loans market, and their payment behaviour from 2018 to 2023, along with macroeconomic data from the "Instituto Nacional de Estadística".

Keywords: financial sector, risk management, collections and recoveries, recovery rate, debt segmentation, modelling, analytics, and COVID-19.

RESUMEN

Las actividades de recobros y recuperaciones son esenciales para una gestión eficaz del riesgo crediticio, particularmente a la luz de las regulaciones posteriores a la crisis financiera de 2007-2008. A pesar de su importancia, esta área a menudo está poco explorada. Una segmentación precisa de la deuda impagada es crucial para optimizar las estrategias de recobro y evaluar su potencial de recuperación. Nuestro objetivo es identificar las variables clave que influyen en las tasas de recuperación de carteras de deuda no garantizadas que los bancos e instituciones financieras han pasado a fallidos, y evaluar cómo estas variables se vieron afectadas por la crisis del COVID-19. Específicamente, buscamos identificar qué variables predicen significativamente el comportamiento de pago a corto plazo antes, durante y después de la pandemia. Para ello, desarrollamos modelos estadísticos utilizando la metodología de árbol de clasificación CHAID para identificar las variables más discriminantes. Nuestros hallazgos indican que la crisis del COVID-19 tuvo un impacto notable, causando alteraciones significativas en estas variables. Se espera que los conocimientos obtenidos de esta investigación sean valiosos para los principales actores en el sector de recuperaciones, incluidas entidades acreedoras, gestores de cobro y reguladores/supervisores. En esta investigación, utilizamos una base de datos de una destacada firma de inversión, que contiene información sobre 900.000 cuentas de carteras de deuda no garantizada provenientes del mercado español de créditos impagados, y su comportamiento de pago entre 2018 y 2023, junto con datos macroeconómicos del Instituto Nacional de Estadística.

Palabras clave: sector financiero, gestión de riesgos, recobros y recuperaciones, ratio de recuperación, segmentación de la deuda, modelos, analítica y COVID-19.

TABLES INDEX

FIGURES INDEX

GLOSSARY OF TERMS

AdaBoost – Adaptive boosting

AI – artificial intelligence

ANN – artificial neural networks

AR – activity rate

BCBS – Basel Committee on Banking Supervision

BC-OLS – Box-Cox transformation/OLS

BNB – Bernoulli naive bayes

B-OLS – beta transformation/OLS

BR – beta regression

CART – classification and regression tree

C&R – collections and recoveries

CCI – Consumer Confidence Index

CHAID – Chi-square automatic interaction detector

CRPH – Competing risks proportional hazards

DCA – debt collection agency

DNN – deep neural network

DOB – date of birth

DT – decision tree

EAD – exposure at default

EBA – European Banking Authority

ESCB – European System of Central Banks

EL – expected losses

GBDT – gradient boosting decision trees

GCPI – General Consumer Price Index

GDP – Gross Domestic Product

GDP-FTEE – GDP Full-Time Equivalent Employment

GDP-MP – GDP Market Price

GEV – generalised extreme value

GLM – generalized linear model

GNB – Gaussian naive bayes

INE – "Instituto Nacional de Estadística", National Institute of Statistics

IV – information value

IVR – interactive voice response

KNN – K-nearest neighbours

KPI – key performance indicator

LDA – linear discriminant analysis

LGD – loss given default

LR – linear regression

LSSVM – least squares support vector machine

LTV – loan to value

MAR – missing at random

MARS – Multivariate adaptive regression splines

MCAR – missing completely at random

MDA – Multiple discriminant analysis

ML – machine learning

MLP – multilayer perceptron model

MNAR – missing not at random

NN – neural network

NPL – non-performing loan

OLS – ordinary least squares

P2P – peer to peer

P&L – Profit and Losses

PD – probability of default

PLTR – penalised logistic tree regression

PMI – purchase manager index

PTP – promise to pay

QDA – quadratic discriminant analysis

QMLE – quasi-maximum likelihood estimation

QR – quantile regression

QRNN – quantile regression neural network

RiR – Ridge regression

RF – random forest

RoR – Robust regression

RPC – right-party contact

RR – recovery rate

RT – regression tree

S&P – Standard & Poor's index

SMEs – small and medium-sized enterprises

SPV – special purpose vehicles

SVM – support vector machine

SVR – support vector regression

UCPI – Underlying Consumer Price Index

UR – Unemployment rate

XGBoost – Extreme gradient boosting

WHO – World Health Organization

W-O – written-off

WOE – weight of evidence

GRAPHICAL SYNOPSIS

Introduction

Thesis objective

↓

Literature review

- Debt C&R activities and industry
- LGD and RR research
- Segmentation relevance
- COVID-19 effect

↓

Data treatment and analysis

- Data cleaning and normalization
- Variable to be explained definition
- Explanatory variables construction

↓

Models construction

- Methodology definition
- Constructed models: Individuals and SMEs / Global, Pre-COVID, COVID and Post-COVID

↓

Conclusions and recommendations

SECTION 1.- INTRODUCTION

1.1.- OBJECTIVES

1.2.- METHODOLOGY USED IN THE RESEARCH

1.3.- THESIS ORGANIZATION

1.1.- OBJECTIVES

The primary objective of this research is to deepen the understanding of debt segmentation and the influence of crises, such as the COVID-19 pandemic, on this field. Specifically, it seeks to identify the key determinants of recovery rates for unsecured debt portfolios that banks and financial institutions have written off, and to evaluate how these determinants have been affected by the COVID-19 crisis (and all the measures delivered to alleviate the pandemic impact): which variables significantly predict short-term repayment behaviour before, during, and after the COVID-19 crisis?

The adage "past collections do not guarantee future collections" is a well-known principle within the collections and recoveries (C&R) industry. However, analysing and comprehending past collection efforts can significantly enhance future collections by facilitating effective debt segmentation and the development of customized collections and recovery strategies tailored to each identified segment (Crespo & Govindarajan, 2018; Higginson et al., 2018).

Segmentation of debt is applied across all the different management phases (preventive, early arrears, late arrears, and written-off) for a better understanding of the debt and the debtor characteristics. Debtor profiles exhibit significant variability, encompassing categories such as prompt payers, slow payers, careless debtors, recalcitrant debtors, liar individuals, deadbeat debtors, skip debtors, and professional credit offenders, among others. A nuanced understanding of these distinct debtor profiles is imperative for categorizing delinquent accounts, assessing their collection potential, and formulating tailored strategies to optimize collections within each segment (Coleman, 2004). The details of these strategies include, among others, the type and number of amicable actions (e.g., calls, SMSs, or letters), the initiation of litigation, and the prioritization of accounts based on estimated collectability.

The financial crisis of 2008 precipitated a substantial escalation in the volume of non-performing assets, defined as financial exposures in a state of insolvency. This scenario underscored the necessity for rigorous modelling of the various facets of credit risk to facilitate sound risk management practices (Várgedő, 2024; Carleo & Rocci, 2024; Ramos Gonzalez et al., 2023; Gambetti et al., 2019). The Basel II and III Accords established essential credit risk parameters for estimating expected losses (EL), including the probability of default (PD), exposure at default (EAD), loss given default (LGD), and recovery rate (RR). It is noteworthy that LGD can be computed using the formula LGD = 1 - RR (Nazemi & Fabozzi, 2024; Tanoue et al., 2017). The recovery rate signifies the proportion of funds that financial institutions successfully reclaim from non-performing loans (NPLs) (Carleo & Rocci, 2024). As a result, C&R activities significantly influence primary credit risk parameters by affecting LGD and RR.

One notable challenge confronting the banking sector has been the proliferation of unpaid debts, which has spurred the emergence of the debt collection industry and the practice of debt purchasing. Debt collection agencies (DCAs) have emerged as pioneers and specialists in this domain (Shoghi, 2019).

The onset of the COVID-19 crisis at the close of 2019, culminating in global repercussions throughout 2020, engendered a profound health, social, and economic crisis (Lin & Lu, 2024). In response to the pandemic, a range of measures was enacted to curb virus transmission and alleviate its economic and social repercussions. These measures, which included restrictions on movement and business operations, contributed to a severe economic recession that affected nearly every sector and activity. Credit risk management, along with C&R, faced substantial challenges during this period (Lin & Lu, 2024; Ramos Gonzalez et al., 2023; O'Brien et al., 2021; Altman, 2021).

Although references to unpaid debts and C&R activities have long been prevalent in the general literature, scholarly research in this area has been relatively scarce (Malinconico & Virguti, 2024; Kriebel & Yam, 2020; Cheng & Cirillo, 2018; Thomas et al., 2016). However, interest in this subject has increased markedly in recent years (Li et al., 2023A; Sopitpongstorn et al., 2021).

Nearly two centuries ago, Balzac dedicated a novel to portraying what could be conceptualized as a manual for the "professional" debtor, illustrating methods of subsisting without repaying debts (Balzac, 1827). In stark contrast to the characters depicted in Balzac's works, C&R activities allocate considerable resources and efforts towards mitigating the repercussions of unpaid debts across various economic sectors linked to credit. In doing so, they help foster a healthier economic environment (ECB, 2017; Drozd & Serrano-Padial, 2013).

The present research posits several additional and complementary objectives. First, it aims to enhance understanding of how crises, particularly those triggered by COVID-19, have influenced the C&R sector, specifically regarding the criteria used in debtor segmentation. Secondly, it seeks to disseminate these insights among professionals within the debt management field. Lastly, it aspires to bridge the existing gap between the C&R sector and academic inquiry.

To facilitate the achievement of these objectives, access to a database from a preeminent Investment Firm within the Spanish NPL market has been granted. This database comprises detailed information on unsecured debt portfolios sourced in Spain, with an average debt duration of 11 years for Individuals and 12 years for small and medium-sized enterprises (SMEs), along with pertinent account characteristics and payment behaviours documented from 2018 to 2023. The aforementioned database is readily accessible for consultation by the doctoral tribunal, should the need arise.

1.2.- METHODOLOGY USED IN THE RESEARCH

This thesis has been developed in five main phases. First, a review of the academic literature and best practices in C&R was conducted. Next, data gathering, processing, and analysis were conducted. The third and fourth phases involved developing statistical models and reviewing them with practitioners, respectively. Finally, the results were analysed, and conclusions were drawn.

Phase 1. Review of Academic Literature and Best Practices in C&R

This research began with a comprehensive bibliographic review of the C&R industry, with a specific focus on RR and LGD modelling. The goal was to identify the primary determinants of debt segmentation and to analyse the most widely used and effective modelling methodologies. Additionally, this academic review was compared with the methodologies and determinants used by practitioners in C&R best practices.

In the case of modelling methodologies, there is alignment between academic literature and standard C&R practices. The most commonly used methodologies for debt segmentation by practitioners are also represented in the academic literature. Regarding the main determinants, the variables identified in both literature and practice were complementary. As a result, a comprehensive list was created that incorporated variables from both sources.

Phase 2. Data Gathering, Treatment, and Analysis

After selecting the variables, data were collected from two sources: the Investment Firm systems and the "Instituto Nacional de Estadística" (INE) public information website.

The data from the Investment Firm systems included information about account and debtor characteristics, payment behaviour from 2018 to 2023, and the collection activities conducted during that period. Meanwhile, several macroeconomic factors were extracted from the INE website and added to the research dataset.

Prior to modelling, data cleaning and normalization were performed to improve analysis efficiency and derive meaningful conclusions from the dataset.

Phase 3. Development of Statistical Models

In this phase, we undertook several activities to develop our statistical models. First, we defined and created the dependent variable: the probability that an account will make a payment within the next 30 days. Additionally, we created derived variables that could add value to the modelling process using the original dataset variables. Following this, we conducted univariate and bivariate analyses of the dataset's variables to assess their correlation with the dependent variable and their relevance to the models.

The global dataset was divided into two separate datasets: one for Individuals and another for SMEs. This division was based on the differences in payment behaviour and modelling identified through dataset analysis, academic literature (Dumitrescu et al., 2022), and C&R practices. Statistical models were developed using the CHAID classification tree methodology to predict the defined variable for both Individuals and SMEs. The analysis covered the entire dataset period from 2018 to 2023, as well as three specific timeframes: the pre-COVID-19 period (2018-2019), the COVID-19 period (2020-2021), and the post-COVID-19 period (2022-2023). This approach followed a similar timing split as noted in the academic literature (Xu et al., 2024). In total, eight models were created.

The models were initially developed using the full range of variables available in the dataset, which included clients who had previously made payments (payers). However, due to the high correlation observed with prior payments, the models were revised to exclude the payers. Finally, a more streamlined version of the models was created, incorporating only the most relevant variables. All models were thoroughly analysed to assess their predictive quality, strength, and stability, utilizing standard market key performance indicators (KPIs) such as the Gini coefficient and the Kolmogorov-Smirnov test.

Phase 4. Review of the Models with Practitioners

The models developed were shared with several practitioners to gather their feedback on the results obtained and how they compared to their own experiences. Based on this feedback, we made several adjustments to enhance the models.

Phase 5. Analysis of Results and Conclusions

In this final phase of the research, we analysed and compared all models to assess whether the main determinants of segmenting debt portfolios were consistent across the Global, Pre-COVID-19, COVID-19, and Post-COVID-19 models.

1.3.- THESIS ORGANIZATION

The structure of the present thesis aligns with the objectives and methodological phases previously outlined for this research. The first section of the document outlines the aims of the study, the methodology used to develop it, and the organization of the thesis.

Section 2 contextualizes our research through an overview of the credit market and the C&R activities, provides a summary of the literature review relevant to the research area, and highlights the identified research gap. This bibliographic review focuses explicitly on the C&R business and the modelling of RR and LGD to identify the main determinants of debt segmentation and to evaluate the most widely used methodologies and those that yield the best performance. Additionally, it contrasts these methodologies and determinants with those employed by practitioners in the C&R sector.

In Section 3, we present the data used in this research, along with the data-cleaning and normalization processes. This section also describes the dependent variable the models aim to explain: the likelihood that an account will make a payment within the next 30 days. Additionally, we outline the final datasets and key variables used, comprising 81 variables sourced from 24 data files. These files contain information on account characteristics, debtor profiles, payment behaviour from 2018 to 2023, collection activities, and macroeconomic factors.

Section 4 focuses on the methodology employed in the research, detailing the modelling framework, the statistical methods applied, and the final models developed. Eight models were constructed using the CHAID classification tree methodology, distinguishing between Individuals and SMEs, and between the Overall period (2018-2023), the Pre-COVID-19 period (2018-2019), the COVID-19 period (2020-2021), and the Post-COVID-19 period (2022-2023).

The subsequent Section 5 presents the primary conclusions and recommendations drawn from this research, along with additional research areas that may emerge from the current thesis.

Additionally, Section 6 includes the references used throughout the development of this study, while Section 7 features appendices that provide details on the data files, data cleaning and normalization processes, the explanatory variables employed, the macroeconomic factors considered, the models developed in this research, and a summary of the present thesis.

SECTION 2.- LITERATURE REVIEW AND RESEARCH GAP

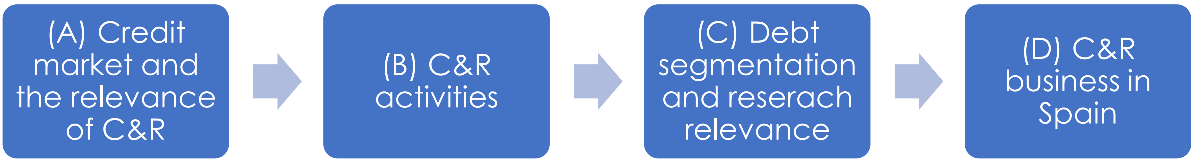
2.1.- INTRODUCTION

2.2.- LITERATURE REVIEW AND RESEARCH GAP

2.1 INTRODUCTION

In this subsection, we aim to contextualize our research by providing an introductory overview that encompasses: (A) the credit market and the integration of C&R activities within this framework; (B) a detailed examination of the specific C&R activities; (C) a review of the debt segmentation and its significance; and (D) an exploration of the C&R business landscape in Spain.

Figure 1. Context of the research



Source: Own elaboration

(A) Credit Market and the Relevance of Collections and Recoveries

The relevance of the credit market and C&R is significant. Credit refers to the provision of a benefit—whether in the form of a product, service, or monetary amount—with deferred payment. This practice is prevalent across sectors of the global economy, including banking, finance, retail, telecommunications, and utilities. When credit is extended for the delivery of products and services, there is inevitably a risk of non-payment, commonly referred to as credit risk. This type of risk is among the oldest in financial markets (Molina et al., 2015; Spuchľakova & Cug, 2015; Stifler & Parrish, 2014).

The primary phases of the credit risk management cycle can be articulated as three key components (Molina et al., 2015): Admission – this phase involves determining the customers eligible for the product or service, along with the specific conditions under which it will be granted, including factors such as pricing, loan amounts, instalment terms, and required guarantees; Portfolio management – following the approval of the product or service, this step entails establishing the controls and policies necessary for effective credit risk management; and Collections & recoveries (C&R) – in this phase, policies and procedures are put in place to address unpaid debts and manage situations of default.

The banking and finance sector is the most advanced and sophisticated area of credit risk management, primarily due to its inherent nature and critical significance. In comparison, sectors such as telecommunications and utilities are less developed in this regard. This distinction likely explains the greater breadth of regulations, research, and publications in banking and finance than in other industries. Consequently, our references in this research will predominantly draw from this sector.

The financial crisis of 2007-2008 prompted regulators to intensify efforts to strengthen banks' and financial institutions' risk management practices, particularly in the computation and reporting of credit losses. This led to enhancements in the Basel Accords' Internal Ratings-Based (IRB) approach for modelling credit losses and its implications for credit risk management (Sopitpongstorn et al., 2021; Gambetti et al., 2019).

The Basel Accords have played a pivotal role as a regulatory framework for banks and financial institutions over the past 35 years. The Basel Committee on Banking Supervision (BCBS) comprises central banks and regulatory authorities responsible for banking oversight from key economies worldwide. As of 2024, the committee includes 45 members from 28 jurisdictions. Through the Basel Accords, the BCBS develops guidelines for banking regulation, offering insights and implications for effective risk management in the industry.

In the year 2000, the BCBS released a set of principles for managing credit risk, known as the Credit Risk Principles. These principles aim to encourage banking supervisors worldwide to advocate for sound credit risk management practices. Over the years, the specifics of these principles have been periodically updated (BCBS, 2025; BCBS, 2000).

Several of these principles, whether directly or indirectly, highlight the significance of C&R activities within the broader framework of credit risk management.

In the section titled "Operating under a sound credit-granting process", Principle 4 emphasizes that a thorough understanding of the borrower's or counterparty's risk profile and characteristics must be integral to a sound and clearly defined credit-granting process. Additionally, the borrower's repayment history and repayment capacity must be considered among the key factors.

Within the section "Maintaining an appropriate credit administration, measurement, and monitoring process", Principle 9 stipulates that banks must implement policies, processes, and methodologies to effectively grade, classify, and monitor all credit exposures, including off-balance-sheet and forborne exposures. Furthermore, Principle 11 asserts that banks must use information systems and analytical techniques to enable management to accurately measure the credit risk inherent in all on- and off-balance-sheet activities.

In the section titled "Ensuring Adequate Controls Over Credit Risk" within the Credit Risk Principles, Principle 15 stipulates that banks must implement a system for promptly addressing deteriorating credits and managing problem exposures, as well as related workout situations. Effective workout programs are essential for effectively managing risk within the portfolio.

The Basel II Accord established the framework that allows banking and financial institutions to use their internal credit risk parameters to calculate expected losses and capital requirements. This methodology is known as the internal ratings-based (IRB) approach. The three components of expected losses (EL) under the IRB approach are as follows: Probability of Default (PD) – this component indicates the percentage of borrowers expected to default within one year; Exposure at Default (EAD) – this metric estimates the amount that remains outstanding in the event of a borrower's default; and

Loss Given Default (LGD) – this shows the percentage of exposure that the entity might lose if a borrower defaults.

The equation for expected losses is given by: EL = PD * EAD * LGD

Typically, LGD is expressed as LGD = 1 - Recovery Rate (RR), where the Recovery Rate is defined as the proportion of the amount outstanding that is successfully collected or recovered by the entities, after deducting administrative fees incurred during the collection process (Galow et al., 2024; Hoechstoetter & Nazemi, 2013; Repullo et al., 2010; BCBS, 2005).

Banks that employ LGD models with strong predictive capabilities can gain a significant competitive advantage, whereas models with poor predictive performance may lead to adverse selection (Gürtler & Hibbeln, 2013).

As illustrated below, the activities involved in collecting and recovering outstanding amounts are crucial to LGD, underscoring the vital role of C&R management in credit risk management.

In recent decades, increased competition among banks has influenced their credit risk, leading to a rise in non-performing loans (NPLs) (Radivojević et al., 2019). NPLs are a significant concern for banks, regulators, and market participants, as they pose risks to the overall economy and financial system. These loans negatively impact banks' profitability, consume valuable resources, and may restrict banks' ability to lend to the real economy (EBA, 2019). As a result, NPLs have garnered substantial attention from European regulators in recent years. Among their guidelines, regulators recommend that banks consolidate their NPLs and sell them to specialized investors, thereby mitigating impairment losses and addressing financial stability concerns (Bellotti et al., 2021).

Debt buyers and investors are specialized firms that acquire charged-off or delinquent debts from banks, credit card companies, and other creditors for a fraction of the total amount owed. After obtaining these debt portfolios, they strive to collect outstanding debts and, in some instances, may even sell the debts to other buyers. This practice of debt buying represents one of the most significant developments in the debt collection industry in recent years. Key components of this market include debt buyers, debt sellers, the pricing of debts—which is influenced by factors such as demand, availability, age of the debt, frequency of prior collections or sales, geographic location, and documentation—and the agreements negotiated between buyers and sellers. These agreements define the price and face value of the debts included in the sale, the types of debts in the portfolio, the accompanying information, the accuracy of the account data, and any supporting documentation for the accounts (Stifler & Parrish, 2014).

The European market for debt collection and management services originated in the Nordic countries during the 1990s, driven largely by activities in the real estate and banking sectors. Several of the largest European debt collection agencies and investors, such as Intrum and Axactor, have their headquarters in the Nordics and maintain branches across Europe. The financial crisis of 2008-2010 spurred the growth of the servicers industry in the UK, Spain, and Italy, partially influenced by directives from European regulators. In the wake of the crisis, the NPL ratio increased notably, prompting the European Banking Authority (EBA) to issue guidelines and a Council Action Plan to reduce distressed receivables and prevent the accumulation of new distressed assets. This action plan proposed measures in three main areas: enhancing supervision; developing secondary markets, with key strategies like debt portfolio sales and securitizations to drive NPL reduction; and reforming insolvency frameworks (Malinconico & Virguti, 2024).

(B) Collections and Recovery Activities

When a debtor fails to meet the repayment obligations under the terms of a credit agreement, the debt enters a delinquent phase and is subsequently classified as an NPL (Malinconico & Virguti, 2024). To collect and recover outstanding amounts, creditors engage in time-consuming, costly processes. The primary and most common activities undertaken fall into three categories: amicable management, judicial management, and debt sale (Malinconico & Virguti, 2024; Shoghi, 2019; Stifler & Parrish, 2014; Bailey, 2002).

Amicable management involves friendly measures that creditors take to reach out to debtors and negotiate debt repayment. These measures typically include phone calls, text messages, letters, emails, and personal visits by field collectors (Malinconico & Virguti, 2024; Shoghi, 2019; Thomas et al., 2012; Rosenberg & Gleit, 1994). Recently, C&R practices have evolved to incorporate various digital channels, including interactive voice response (IVR), virtual agents, chatbots, and online banking applications, for notifications and communication.

In the context of judicial management, the primary actions involve legal processes to recover debts. Furthermore, it is common practice in C&R to conduct these legal procedures alongside amicable efforts. This dual approach not only strengthens negotiations with debtors but also improves the overall success of debt collection (Malinconico & Virguti, 2024; Stifler & Parrish, 2014; Thomas et al., 2009).

In the final stages of collections and recovery efforts, a creditor may consider selling debts to third-party investors, commonly referred to as a debt sale. This typically occurs during the later phases of the debt lifecycle when there is a lack of anticipated recovery from both amicable and judicial management strategies (Fonseca et al., 2017; Stifler & Parrish, 2014; Hoechstoetter et al., 2012).

Creditors can implement C&R activities using internal resources, referred to as "in-house" C&R, or by outsourcing them to external providers, known as "third-party" C&R. A prevalent approach among creditors is a hybrid model that combines both in-house and third-party resources (Li et al., 2021; Fedaseyeu, 2020; Fedaseyeu & Hunt, 2015; Matuszyk et al., 2010).

C&R management is typically organized into several sequential phases, primarily delineated by the age of the delinquent debt. These phases include preventive measures, early arrears management, late arrears management, and the handling of written-off debts (Malinconico & Virguti, 2024; ECB, 2017; Bailey, 2002; Lawrence, 1992). Each phase is characterized by distinct strategies and interventions to manage the debt recovery process effectively.

Preventive management involves proactive measures for accounts that are not yet delinquent but are identified as at risk of becoming unpaid. This is achieved through a combination of statistical models and expert criteria to prevent these accounts from entering delinquency. Following the preventive management phase, the process continues with the management of early and late arrears. Early arrears management focuses on amicable interventions for accounts in the initial stages of delinquency — typically 1 to 90 or 120 days overdue —and covers cases with 1 to 4 missed debt instalments. Conversely, late arrears management encompasses both amicable and judicial activities for accounts in more advanced stages of delinquency, which generally fall within the 90 to 120-day range and extend to written-off debt. The timing of written-off debt can vary by debt type and country-specific legislation, typically occurring when debts are between 2 and 7 years old. In this phase, the sale of credits is also a standard practice among creditors. Finally, we arrive at the written-off debt management phase, which incorporates both amicable and judicial strategies for accounts that have already been written off—these are the cases that are fully provisioned and no longer reflect on the creditors' current balance.

Regulators and supervisors in the banking industry emphasize that during the C&R phase of the credit risk cycle, optimal and effective management practices should be centred around governance and operating models (ECB, 2017). The ECB outlines key components of sound governance, including strategic objectives, decision-making policies and processes, a robust control framework, monitoring systems, reporting systems, and early warning mechanisms.

Banks must define and implement their NPL management strategy within their overall decision-making policies and processes. This typically involves several key steps: first, conducting an assessment of the operating environment, including evaluating internal capabilities and external factors that impact NPL resolution. The ECB emphasizes several points relevant to our research, such as the importance of understanding the outcomes of NPL management actions, recognizing investor demand for debt portfolios, and being aware of the NPL servicing industry. The next step is to develop the NPL strategy, which should include setting targets for both operational capabilities (qualitative) and planned NPL reductions (quantitative) across short, medium, and long-term time horizons. The final steps involve implementing the operational plan and integrating the NPL strategy into the bank's management processes.

The control framework should encompass the three lines of defence employed by banks. The first line of defence comprises the control mechanisms integrated within operational units that actively manage risks and conduct compliance and risk activities. The second line of defence includes risk control, compliance, and various quality assurance functions, which are designed to ensure, on an ongoing basis, that the first line of defence operates effectively. Lastly, the third line of defence is represented by the internal audit function.

The ECB emphasizes the importance of monitoring, reporting, and early warning systems within the governance of banks' C&R activities. In their monitoring frameworks, banks are expected to incorporate key performance indicators (KPIs) that enable management to assess progress against established targets. These KPIs fall into five distinct categories: 1) high-level NPL metrics, which include the NPL ratio and coverage, NPL flows, default rates, migration rates, and probabilities of default; 2) customer engagement and cash collection metrics, such as scheduled versus actual borrower engagements, the percentage of engagements leading to payments or promises to pay, and total cash collected; 3) forbearance activities, focusing on the efficiency and effectiveness of these measures; 4) liquidation activities, which encompass legal actions, foreclosures, and debt-to-asset/equity swaps; and 5) additional metrics related to profit and loss items, foreclosed assets, early warning indicators, or outsourcing activities.

To effectively manage NPLs, banks should establish a comprehensive reporting system that periodically informs supervisors about NPL monitoring KPIs and any significant structural changes in the NPL operating model or control framework. Moreover, best practices should incorporate early warning mechanisms to monitor performing loans and prevent a decline in credit quality. This includes implementing robust internal procedures and reporting systems to identify and address potential non-performing clients at an early stage.

In terms of the operating model, the ECB emphasizes the specific functions of the C&R units, distinguishing them from the Admission or Loan Origination units. C&R operations must be tailored to the various stages of the debt life cycle, including early arrears (up to 90 days past due), late arrears/restructuring/forbearance, liquidation/debt recovery/legal cases/foreclosure, and foreclosed assets. It is also essential for banks to implement a debt portfolio segmentation strategy that allows them to categorize borrowers with similar characteristics who require similar treatment. Additionally, designing customized processes for each defined segment is critical.

Moreover, adequate human and technical resources must be established. The necessary human resources should be determined based on their proportionality, expertise, experience, and performance management. A robust technical infrastructure is a key factor in successfully executing any NPL strategy. This infrastructure should facilitate easy access to all relevant data and documentation for NPL workout units, enable efficient processing and monitoring of NPL workout activities, and support the definition, analysis, and measurement of NPLs and borrowers, including NPL segmentation analysis.

(C) Debt Segmentation and Research Relevance

As previously noted, and supported by various studies in the literature, debt segmentation is crucial in the C&R business for several reasons: it informs strategies, prioritizes activities, optimizes resource allocation, defines objectives, and ultimately enhances outcomes (Malinconico & Virguti, 2024; Kriebel & Yam, 2020; Hunt, 2007; Piumelli & Schmidt, 1998). A company's C&R strategies form the core of its collections process, and more effective debt segmentation leads to improved strategies and better results (Piumelli & Schmidt, 1998).

Higginson et al. outline seven essential pillars for best-in-class collections operations, with segmentation being the foremost. The remaining six pillars include skip tracing, contact strategy, treatment choices, frontline capabilities, performance management, and organization. The authors emphasize that the primary purpose of segmentation is to differentiate between delinquent customers who need human contact and those who can be effectively engaged through automated messaging or may require no contact at all. Furthermore, they assert that integrating machine learning and advanced analytics will enhance both value-at-risk and behavioural segmentation, thereby refining the contact and treatment strategies tailored to each defined client segment (Higginson et al., 2018).

Crespo & Govindarajan emphasized that the most significant advancements in collections are being driven by advanced analytics and machine learning. These innovations are revolutionizing collections operations, enhancing performance at a more granular level and enabling more sophisticated, productive customer segmentation. This improved segmentation forms the foundation for more effective collections processes and strategies, enhancing the entire collections cycle—from preventing and managing bad debt to facilitating internal and external account resolutions. By leveraging analytics algorithms, organizations can gain a deeper, more nuanced understanding of at-risk customers, classify them into microsegments, and design more targeted and effective interventions. Some tangible benefits of this approach include early identification of self-cure customers, leading to considerable savings in collectors' time and resources—evidence shows that some banks have increased collector capacity by 5-10% through more sophisticated self-cure models. Additionally, this method prioritizes value-at-risk over standardized criteria, enabling entities to reduce charge-offs by 5-15%. Moreover, more clearly defined management paths between cure or pre-charge-off offers, such as early settlements, have been observed. By optimizing pre-charge-off offers, there is potential to reduce charge-offs by 10-20% through early settlement options. Lastly, enhanced post-charge-off management decisions, particularly regarding assignment to third-party agencies or pricing segmentation, can further improve outcomes (Crespo & Govindarajan, 2018).

This research examines debt segmentation, specifically analysing the critical variables that influence it in unsecured debt portfolios written off and acquired from banks and financial institutions. Furthermore, it investigates the impact of the COVID-19 pandemic (and the measures implemented to mitigate its effects) on these variables and their role in shaping debt segmentation outcomes.

The COVID-19 pandemic and its widespread implications have significantly affected the social and economic landscape, both globally and within Europe and Spain. It has heightened market uncertainty, disrupted companies' production and operations, and affected individuals' incomes. This, in turn, has diminished their capacity and willingness to repay debts, thereby increasing the risk of loan defaults. As a result of these effects, various sectors of economic activity have been compelled to evolve and completely rethink their management strategies, particularly in operational and analytical areas. Credit and risk management, along with C&R—a vital aspect of it—are no exceptions. Since the onset of the pandemic, these processes have undergone reinvention (Malikkidou & Strohbach, 2025; Xu et al., 2024; Lin & Lu, 2024; Laurent et al., 2020).

A recent survey conducted by SAS and FT Longitude, which included 300 banks worldwide, revealed that 67% of the banks intend to modernize their risk modelling practices within the next two years. This initiative is deemed essential, with 63% of respondents indicating it could provide a competitive advantage. Additionally, artificial intelligence (AI), machine learning, and data analytics tools have emerged as the primary investment priorities for risk modelling among the surveyed banks (SAS, 2025).

The findings of this research are likely to be significant for the key stakeholders in the C&R sector, including creditors, servicers, and regulators/supervisors. A deeper understanding of how a crisis such as the one brought about by COVID-19 impacts debt segmentation could lead to improved modelling, enhanced preparation for economic downturns, more effective regulation, refined processes, and ultimately, better outcomes within the C&R landscape.

(D) Collections and Recoveries Business in Spain

To effectively contextualize the Spanish market, it is essential to examine the significance of the debt C&R sector by analysing data from the annual market reports published by DBK Informa, ANGECO, and Axis Corporate. This examination will facilitate a comprehensive understanding of the dynamics and relevance of the debt C&R business within the Spanish economic landscape.

DBK Informa emphasizes the significant atomization of the Spanish market, which is characterized by the presence of over 800 companies and a notable influence from foreign firms (Informa, 2023). In its annual report for the end of 2022, similar figures to those reported by ANGECO are noted:

- Approximately €300 million in debt under management
- €1.8 billion collected, representing 4% of the total debt
- €1.5 billion in total revenues
- Around 20,000 employees
- €14.6 billion in debt portfolios sold

ANGECO, the National Association of Collection and Recovery Entities— "Asociación Nacional de Entidades de Gestión de Cobro" — publishes its Annual Market Study, which compiles key market figures from its members (ANGECO, 2024). In 2023, entities affiliated with ANGECO accounted for 85% of the debt collection and recovery sector in Spain, with the following key indicators:

- 124 million files managed
- €463,973 million in debt under management
- €12,422 million collected
- €1,748 million in total revenues, with €1,419 million derived from C&R activities

- A total workforce of 21,086 employees, of which 17,176 are involved in C&R

To establish a reference point for the sector's significance, it is noteworthy that the revenue generated by ANGECO associates in 2023 accounted for nearly 1% of Spain's Gross Domestic Product (GDP) that year. This statistic underscores the sector's substantial contribution to Spain's overall economic landscape.

Axis Corporate analyses the Spanish market in its latest annual report (Axis Corporate, 2025). The report highlights the ongoing concentration and reconfiguration process within the C&R servicing sector, emphasizing the integration of AI across all of its activities. Furthermore, it includes a survey of key industry players, revealing that many regard AI and advanced data analysis as the most significant factors impacting their portfolio management success.

Additionally, Axis Corporate estimates that the volume of debt portfolios sold in the Spanish market will reach €17 billion in 2024, with an overall annual market size projected at approximately €15 billion.

2.2 LITERATURE REVIEW AND RESEARCH GAP

The objective of this research is to identify the primary determinants of RRs for unsecured debt written-off portfolios acquired from banks and financial institutions, as well as to assess the impact of the COVID-19 pandemic (and all measures delivered to mitigate its effects) on these determinants: which variables significantly predict short-term repayment behaviour before, during, and after the COVID-19 crisis? We commenced the study with a literature review focused on the C&R business, alongside modelling methodologies for RRs and LGD. This review aimed to clarify the key factors influencing debt segmentation and to evaluate the most widely used and practical modelling approaches. Our main findings were then compared with the methodologies and determinants used by practitioners in C&R best practices.

As previously stated in this document, the primary phases of the credit risk management cycle include Admission, Portfolio Management, and Collections & Recoveries (Molina et al., 2015). In practical applications of credit risk management, there has historically been a predominant focus on the admission phase, as this area is most pertinent to the business's economic viability. This trend is also reflected in the academic literature, which emphasizes the Probability of Default (PD) associated with the admission phase. At the same time, LGD, RR, and C&R have received comparatively less attention, despite being critical components of credit risk and essential for effective risk management processes. However, in recent years, there has been a noticeable increase in research concerning LGD, particularly related to the C&R phase, spurred by banking regulations introduced under Basel II and III (Várgedő, 2024; Jacobs, 2024; Loterman et al., 2012; Altman et al., 2004).

In 2015, Livshits presented a comprehensive and insightful literature review on recent advancements in the quantitative analysis of unsecured consumer debt and default. This review addressed key topics, including bankruptcies, the significance of asymmetric information and information technology, debt collection, and the restructuring of distressed debt. It also explored the cyclical behaviour of consumer debt and default, the insurance role of household debt, and the implications for welfare analysis and policy design. Since then, the development of literature in this area has continued to evolve (Livshits, 2015).

In his review of delinquencies, defaults, and collections, Livshits emphasized that most of the papers discussed thus far have concentrated on formal bankruptcy. He noted that, in addition to formal defaults, there is a growing body of literature addressing various stages of delinquency, particularly focusing on collection efforts, self-cure models, debt restructuring, and informal bankruptcies. These insights from Livshits have been incorporated into our literature review.

Based on the findings presented in the literature review, the analysis has been organized into four distinct subsections, which will be further elaborated upon in the subsequent paragraphs:

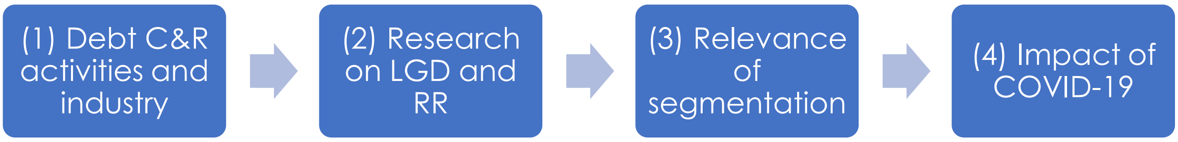
(1) Debt Collection and Recovery Activities and Industry

(2) Research on Loss Given Default (LGD) and Recovery Rate (RR)

(3) The Relevance of Segmentation

(4) The Impact of COVID-19

Figure 2. Literature review structure



Source: Own elaboration

This structured approach facilitates a comprehensive exploration of each topic, thereby contributing to a nuanced understanding of the subject matter.

(1) Debt C&R Activities and Industry

As previously mentioned in this document, C&R activities play a crucial role within the credit risk management cycle, as the cost and effectiveness of the collections process directly impact the pricing and availability of consumer credit. Despite this significance, the collections industry has garnered surprisingly little attention from economic scholars. Generally, the key aspects identified in academic literature regarding the collection of bad debt can be categorized into three main areas: locating debtors, segmentation (which involves distinguishing between consumers who cannot pay due to a lack of resources and those who will not pay even if they have the means), and the use of technology (Malinconico & Virguti, 2024; Fonseca et al., 2017; Thomas et al., 2016; Hunt, 2007). Additionally, analytics-driven intelligence is emphasized as crucial for transforming collections (Crespo & Govindarajan, 2018). The field of collections is recognized as a prime candidate for AI transformation within retail banking, given its high business impact and significant technical feasibility (Giovine et al., 2024). Another important aspect of debt management is minimizing the cost of debt while maintaining an acceptable level of risk—striving for the lowest possible cost with a reasonable degree of prudence. Effective debt management strategies are essential for ensuring financial stability (Jonasson et al., 2019).

The earliest academic work we are aware of that examines debt collections dates back nearly a century and focuses on the debt collection market in the USA (Fedaseyeu & Hunt, 2015; Krumbein, 1924). In this study, Krumbein outlines the operations of debt collection agencies and their market dynamics during that period. This research highlights the separation of collection management from credit management following the

Industrial Revolution and underscores the need to tailor collection strategies and activities to different debtor classes. This concept serves as the foundation for segmenting debts and for defining strategies to optimize C&R outcomes. Furthermore, it discusses how creditors began outsourcing debt management to third-party entities—collection agencies—and the organization and relationship that developed between these agencies and creditors. The primary objectives and requirements for collection agencies are threefold: to achieve the specified collection results, to achieve these results at minimal cost, and to foster amicable relations between creditors and debtors. Internally, collection agencies are structured by dividing their activities into two main functions: 1) client engagement and 2) debtor contact. Creditors contract these agencies based on a compensation model that may consist of a percentage of the recovered funds, a retainer fee for the duration of service, or a combination of both. Such remuneration schemes remain prevalent today (Malinconico & Virguti, 2024; Fedaseyeu, 2015).

Several studies have analysed the impact of regulation on C&R activities and on RRs. Fedaseyeu concluded that stricter laws governing third-party debt collection led to fewer third-party collectors and a decline in RRs for delinquent credit cards (Fedaseyeu, 2020). Similarly, Fonseca et al. found that restrictions on collection practices reduce access to credit and deteriorate indicators of customers' financial health (Fonseca et al., 2017).

Within the academy, we identify three primary approaches or phases for creditors to establish their C&R strategies and processes. These options include managing debt collection internally through in-house departments or teams, outsourcing collection activities to external third-party agencies, or selling the debts. (Carleo & Rocci, 2024; Li et al., 2021; Fonseca et al., 2017; Drozd & Serrano-Padial, 2013).

Delegating the collection of NPLs to external agencies means a bank entrusts these challenging recoveries to specialized companies in exchange for a designated service fee. Third-party debt collectors often employ more aggressive collection practices than the original creditors, and their expertise generally results in lower collection costs or higher recovery rates, as debt collection is their primary focus. Additionally, creditors typically distribute debt collection tasks among various third-party agencies, with each agency frequently representing multiple creditors (Li et al., 2022; Beck et al., 2017; Fedaseyeu & Hunt, 2015).

Debt sales typically occur after a debtor has been delinquent for 120 to 180 days, and after accounts are charged off (Athreya et al., 2018). This practice can be traced back to the 2007-2008 financial crisis, which highlighted the necessity for third-party buyers to relieve originators from the burdensome task of debt collection (Carleo & Rocci, 2024; Hoechstoetter et al., 2012).

Recently, Malinconico & Virguti emphasized that there are currently five distinct business models within debt collection activities, in addition to the original creditors: a) Third-party servicing (Debt collection agencies or third-party servicers): These entities specialize in the collection and management of debt portfolios on behalf of external organizations, such as banks, financial institutions, or investment funds, without having any financial interest in the managed loans; b) Debt acquisition and recovery (Pure servicers): This model focuses on the direct purchase of debt portfolios, followed by the collection and management of these acquired assets; c) Hybrid model of debt acquisition: This approach merges elements of both debt acquisition and third-party servicing, allowing for a flexible strategy in managing debt; d) Pure investors: In this model, investors acquire debt portfolios and delegate the collection and management processes to one of the aforementioned operators; e) Special Purpose Vehicles (SPV): This category is a specific type of investor, where the acquisition of debt portfolios is facilitated through the issuance of typically rated asset-backed securities (Malinconico & Virguti, 2024).

The primary activities involved in collections include refraining from action when payment is anticipated in the coming days, issuing a standard statement that provides a reminder, making telephone calls, sending letters, visiting borrowers in person, and pursuing legal measures. All these efforts aim to achieve one of several outcomes: recovering the full outstanding debt, collecting a portion of the debt through discount offers, establishing a payment plan, initiating legal proceedings, selling the debt, or closing the account (Thomas et al., 2012; Hoechstoetter et al., 2012; Thomas et al., 2009; Rosenberg & Gleit, 1994).

Litigation typically arises when a bank has not effectively communicated with the debtor regarding the NPL or when negotiations and other attempts at resolution have proven unsuccessful. In such cases, the bank initiates a lawsuit to facilitate recovery of the NPL by enforcing a court judgment (Li et al., 2022).

The relevance of information technology (IT) is steadily increasing across various credit management sectors, and C&R is no exception. Numerous studies have underscored this trend. For instance, Chin & Kotak examined the implementation of rule-based technology at Capital One through a case study, demonstrating how it enhanced the bank's collections process. This improvement stemmed from increased information quality, more effective decision-making with customized collection strategies, and greater capacity and flexibility. It is estimated that this evolution in their model has saved Capital One tens of millions of dollars annually (Chin & Kotak, 2006).

Drozd & Serrano-Padial examined the role of information technology in collecting delinquent consumer debt. They concluded that adopting these systems and techniques is essential for lenders to effectively identify clients who are more likely to repay their debts. This approach helps avoid the costs associated with collecting from borrowers who lack the means to repay and enhances the overall efficiency of monitoring and controlling collection expenses. The authors emphasized that the primary distinction between traditional methods and modern collection technologies lies in the systematic application of information and statistical modelling, alongside computer-based optimization techniques to allocate collection resources more effectively (Drozd & Serrano-Padial, 2017; Drozd & Serrano-Padial, 2014; Drozd & Serrano-Padial, 2013)

Hunt highlights the significant role of technological advancements in transforming the collections industry. Key innovations such as diallers, internet technologies, credit scoring systems, information technology, and specialized collections software have emerged as crucial tools. These technologies facilitate the two primary objectives of debt collection: identifying debtors and differentiating between consumers who are unable to pay due to a lack of resources and those who choose not to pay despite having the means (Hunt, 2007).

Shoghi emphasizes the significance of employing AI and machine learning technologies in the collection and analysis of large datasets. He argues that these advanced technologies are crucial for effectively classifying debtors and systematically recommending appropriate actions regularly (Shoghi, 2019).

Recently, Malinconico and Virguti emphasized the significance of data and technology in enhancing the efficiency of collection activities. Tools such as information technology, predictive diallers, speech analytics, internet-based skip tracing, collection scoring, and the integration of data analytics and AI hold significant potential to improve the effectiveness of debt collection operations. According to these authors, modern debt collection technology relies less on litigation and more on the strategic use of information (Malinconico & Virguti, 2024). This shift is clearly reflected in C&R's everyday practices.

(2) Research on LGD and RR

In the context of research on LGD and RR, two primary streams have been identified. The first examines the relationship between Probability of Default (PD) and LGD-RR, while the second focuses on estimation methodologies and the key determinants influencing both LGD and RR (Han & Jang, 2013; Qi & Zhao, 2011; Bastos, 2010).

Altman et al. analysed the evolution of literature regarding the relationship between Probability of Default (PD) and LGD-RR. Their findings indicate that earlier research suggested an inverse relationship between PD and RR. Subsequent studies presented the view that RR were fixed-ratio, rendering PD and RR independent. More recent research, however, demonstrates that PD and RR are negatively correlated (Altman et al., 2005; Altman et al., 2004).

Subsequent research in this area indicates a correlation between default rates, LGD, and RRs. Specifically, a negative correlation is observed for RRs, whereas a positive correlation is observed with LGD (Calabrese, 2012; Bellotti & Crook, 2012). However, some studies report differing results that vary with the stage of the credit cycle, whether during upturns or downturns (Jimenez & Saurina, 2006).

Furthermore, some researchers have expanded their inquiries into the relationship between the probability of default (PD) and LGD-RR while accounting for the macroeconomic environment. During periods of recession or industry downturns, default rates tend to rise while RRs decline. Although there is a stronger correlation between default rates and RRs than with macroeconomic variables, this relationship reveals common dynamic patterns that are associated with the business cycle (Várgedő, 2024; Li et al., 2023A; Bruche & González-Aguado, 2010).

Gambetti et al. reached slightly different conclusions regarding the relationship between PD and RRs, emphasizing that economic uncertainty significantly influences the systematic variations in RR distributions. They observed that with greater economic uncertainty, the mean of the RR distributions tends to decrease, while an increase in default rates corresponds to a higher dispersion in those distributions (Gambetti et al., 2019).

Regarding determinants and methodology, there is no consensus on which are most suitable for modelling LGD and RR. However, various studies shed light on relevant factors and estimation methods (Han & Jang, 2013).

In early literature on credit risk management, it is noted that one advantage of employing quantitative methods is the optimization of collection policies, which can either minimize collection costs or maximize recovery from delinquent accounts (Rosenberg & Gleit, 1994). More recently, LGD modelling has been identified as a valuable tool for adjusting collection processes and strategies (Bijak & Thomas, 2015).

Estimating and forecasting the LGD and RR values present significant challenges (Wu et al., 2024; Jacobs, 2024; Kellner et al., 2022; Thomas et al., 2012). The literature identifies several reasons for these difficulties: the complex circumstances of individual borrowers (Li et al., 2023A); the variety of influencing variables, the prolonged duration of the recovery process, and the diverse nature of the distributions involved (Li et al., 2022); limited data availability, particularly from third-party purchasers, repayment rates typically exhibit a bimodal distribution, with high probabilities clustered near 1 and 0, while collection workout costs remain undisclosed (Nazemi et al., 2022; Bastos, 2010; Grunert & Weber, 2005); the distributions of values often display high irregularity (Li et al., 2021); the RR itself is a continuous variable that is fractional and bounded within [0,1], characterized by bimodal and asymmetric density, necessitating value transformation for statistical purposes (Sopitpongstorn et al., 2021); and finally, effective modelling requires additional data, the application of more advanced techniques, consideration of a greater number of influencing factors, and unfortunately, the current data availability is limited (Chalupka & Kopecsni, 2008).

The challenges in estimating LGD may have led to methodologies for estimating LGD being less advanced than those for PD modelling. Additionally, RR might be the most overlooked factor in credit risk evaluation (Várgedő, 2024; Gambetti et al., 2019). It is common practice in many applications to treat LGD and RR as constant values (Xia et al., 2021; Gambetti et al., 2019; Bruche & González-Aguado, 2010; Matuszyk et al., 2010).

A consensus on the most effective methodology has yet to be established (Li et al., 2023A). Historically, various regression techniques have been employed, with ordinary least squares (OLS) regression serving as the benchmark (Li et al., 2023A; Ye & Bellotti, 2019; Bellotti & Crook, 2012).

Logistic regression continues to serve as the benchmark and standard method in the credit risk industry, primarily due to its simplicity and inherent interpretability. In contrast, many machine learning and ensemble methods—such as those based on decision trees, like random forests—often lack the same level of explainability and interpretability. As a result, most banks still rely on logistic regression models, which offer the significant advantage of straightforward interpretation (Li et al., 2023A; Dumitrescu et al., 2022).

The primary methodological approaches include various linear and non-linear regression methods, applied at one, two, or multiple stages. However, no consensus has emerged regarding which method performs best, as findings differ across various academic studies. A selection of commonly used methodologies from the literature is presented in Table 1.

Table 1. Summary of main LGD and RR model methodologies in literature

Models methodologies	Authors
Linear regression (LR)	Li et al., 2023B; Dumitrescu et al., 2022; Ye & Bellotti, 2019; Nazemi & Fabozzi, 2018; Thomas & Bijak, 2015; Bijak, 2014; Hoechstoetter & Nazemi, 2013; Thomas et al., 2012; Calabrese, 2012; Zhang & Thomas, 2012; Thomas et al., 2009; Chalupka & Kopecsni, 2008
Generalized linear model (GLM)	Beck et al., 2017; Han & Jang, 2013
Logistic regression	Xu et al., 2024; Dumitrescu et al., 2022; Calabrese & Zanin, 2022; Ye & Bellotti, 2019; Beck et al., 2017; Tanoue et al., 2017; Thomas & Bijak, 2015; Serrano-Cinca et al., 2015; Bijak, 2014; Hoechstoetter & Nazemi, 2013; Loterman et al., 2012; Thomas et al., 2012; Calabrese, 2012; Hoechstoetter et al., 2012; Matuszyk et al., 2010; Thomas et al., 2009
Ordinary least squares (OLS)	Jacobs, 2024; Nazemi et al., 2022; Li et al., 2022; Bellotti et al., 2021; Xia et al., 2021; Li et al., 2021; Tanoue et al., 2017; Han & Jang, 2013; Hoechstoetter & Nazemi, 2013; Loterman et al., 2012; Bellotti & Crook, 2012; Khieu et al., 2012; Qi & Zhao, 2011
Logit-transformed linear regression, Logit-beta regression, and Local logit	Li et al., 2023A; Li et al., 2022; Sopitpongstorn et al., 2021; Thomas et al., 2009
Fractional response and logit regression	Kellner et al., 2022; Sopitpongstorn et al., 2021; Hoechstoetter & Nazemi, 2013; Calabrese, 2012; Bastos, 2010; Qi & Zhao, 2011; Chalupka & Kopecsni, 2008
Beta regression (BR), Inflated beta regression, Joint beta regression, Beta mixture model, and Beta transformation/OLS (B-OLS)	Li et al., 2023A; Calabrese & Zanin, 2022; Kellner et al., 2022; Ye & Bellotti, 2019; Tanoue et al., 2017; Hoechstoetter & Nazemi, 2013; Loterman et al., 2012; Thomas et al., 2012; Calabrese, 2012; Thomas et al., 2009
Box-Cox transformation/OLS (BC-OLS)	Hoechstoetter & Nazemi, 2013; Thomas et al., 2009
Ridge regression (RiR)	Li et al., 2022; Bellotti et al., 2021; Xia et al., 2021; Nazemi & Fabozzi, 2018; Hoechstoetter & Nazemi, 2013; Loterman et al., 2012
Robust regression (RoR)	Loterman et al., 2012
LASSO regression	Li et al., 2022; Bellotti et al., 2021; Li et al., 2021; Nazemi & Fabozzi, 2018
Tobit regression	Li et al., 2023A; Tanoue et al., 2017; Thomas & Bijak, 2015; Bijak, 2014; Hoechstoetter & Nazemi, 2013; Bellotti & Crook, 2012
Principal components regression	Nazemi & Fabozzi, 2018
Quantile regression (QR)	Kellner et al., 2022
Multivariate adaptive regression splines (MARS)	Bellotti et al., 2021; Hoechstoetter & Nazemi, 2013; Loterman et al., 2012

Support vector machine (SVM)	Xu et al., 2024; Li et al., 2023B; Dumitrescu et al., 2022; Bellotti et al., 2021; Hoechstoetter & Nazemi, 2013; Hoechstoetter et al., 2012; Chen et al., 2009
Least squares support vector machine (LSSVM)	Nazemi et al., 2022; Hoechstoetter & Nazemi, 2013; Loterman et al., 2012
Support vector regression (SVR)	Nazemi et al., 2022; Xia et al., 2021; Li et al., 2021; Nazemi & Fabozzi, 2018
Survival analysis	Thomas & Bijak, 2015; Serrano-Cinca et al., 2015; Bijak, 2014; Hoechstoetter & Nazemi, 2013; Zhang & Thomas, 2012
Linear discriminant analysis (LDA) and Quadratic discriminant analysis (QDA)	Xu et al., 2024
Multiple discriminant analysis (MDA)	Altman, 1968
WOE approach	Thomas et al., 2009
Generalised extreme value (GEV)	Calabrese & Zanin, 2022
Quasi-maximum likelihood estimation (QMLE)	Khieu et al., 2012
Decision tree (DT)	Li et al., 2024; Li et al., 2023A; Li et al., 2021; Thomas & Bijak, 2015; Bijak, 2014; Hoechstoetter & Nazemi, 2013; Bellotti & Crook, 2012: Matuszyk et al., 2010
Classification and regression trees (CART)	Xu et al., 2024; Nazemi et al., 2022; Kellner et al., 2022; Bellotti et al., 2021; Sopitpongstorn et al., 2021; Xia et al., 2021; Nazemi & Fabozzi, 2018; Thomas & Bijak, 2015; Bijak, 2014; Hoechstoetter & Nazemi, 2013; Loterman et al., 2012; Hoechstoetter et al., 2012; Qi & Zhao, 2011; Bastos, 2010
Penalised logistic tree regression (PLTR)	Dumitrescu et al., 2022
Conditional inference trees	Bellotti et al., 2021
Chi-square automatic interaction detector (CHAID)	Hoechstoetter et al., 2012
K-nearest neighbours (KNN)	Xu et al., 2024; Bellotti et al., 2021; Hoechstoetter et al., 2012
Artificial neural networks (ANN)	Li et al., 2024; Li et al., 2023B; Nazemi et al., 2022; Dumitrescu et al., 2022; Kellner et al., 2022; Bellotti et al., 2021; Sopitpongstorn et al., 2021; Li et al., 2021; Hoechstoetter & Nazemi, 2013; Loterman et al., 2012; Hoechstoetter et al., 2012; Qi & Zhao, 2011
Deep neural network (DNN)	Nazemi et al., 2022
Quantile regression neural network (QRNN)	Kellner et al., 2022
Dmneural network	Hoechstoetter et al., 2012
Gaussian processes	Calabrese & Zanin, 2022; Kellner et al., 2022; Bellotti et al., 2021; Sopitpongstorn et al., 2021; Hoechstoetter & Nazemi, 2013; Qi & Zhao, 2011

Gaussian naive bayes (GNB) and Bernoulli naive bayes (BNB)	Xu et al., 2024
Random forest (RF)	Xu et al., 2024; Li et al., 2024; Li et al., 2023B; Dumitrescu et al., 2022; Li et al., 2022; Bellotti et al., 2021; Xia et al., 2021; Li et al., 2021
Boosting (gradient boosting, gradient boosting decision trees-GBDT, extreme gradient boosting-XGBoost, light gradient boosting-LightGBM, CatBoost, and Adaptive boosting-AdaBoost)	Xu et al., 2024; Li et al., 2024; Nazemi et al., 2022; Li et al., 2022; Bellotti et al., 2021; Xia et al., 2021; Li et al., 2021; Shoghi, 2019; Nazemi & Fabozzi, 2018
Bagging	Bellotti et al., 2021; Nazemi & Fabozzi, 2018
Cubist	Bellotti et al., 2021
Multilayer perceptron model (MLP)	Xu et al., 2024; Xia et al., 2021
Competing risks proportional hazards (CRPH)	Li et al., 2023B

Source: Own elaboration

Decision and regression trees have emerged as compelling alternatives to the parametric models typically used in empirical studies of LGD and RR, such as logistic regression. In some research, these tree-based methodologies have even demonstrated superior performance compared to other approaches (Bastos, 2014; Bastos, 2010; Matuszyk et al., 2010; Rosenberg & Gleit, 1994).

Decision trees were first developed in the early 1960s by H. Raiffa and his colleagues at Harvard Business School. Subsequently, Sparks, Breiman, and Makowski advanced these concepts in the years that followed, particularly applying them to credit scoring. One effective way to utilize a decision tree is to associate each node with the probability of non-payment for the individuals represented by that node. The probability at a given node, denoted as node n, is determined as a weighted average of the probabilities from its child nodes, where the weights correspond to the number of individuals at node n. To decide an observation (or account), we traverse the tree from the root node, selecting the appropriate paths based on the observation until we reach the relevant leaf node. By comparing the probability of non-payment at this node to a predetermined cutoff, we can arrive at a decision. However, a significant drawback of decision trees is the larger sample sizes required to obtain statistically reliable probability estimates at each node. Nevertheless, they offer substantial advantages, such as effectively reflecting the impact of various combinations of factors, allowing for cost-structure variability across nodes, and providing overall ease of use and understanding (Bastos, 2014; Rosenberg & Gleit, 1994).

Dumistrescu et al. introduced a technique that integrates logistic regression with decision trees, known as Penalised Logistic Tree Regression (PLTR). This methodology features a straightforward logistic regression model that incorporates predictors derived from decision trees. PLTR aims to enhance the predictive performance of the logistic regression model through data preprocessing and feature engineering, utilising shallow decision trees and a penalised estimation approach, while maintaining the inherent interpretability of the scoring model. In credit scoring, ensemble methods based on decision trees, such as random forests, achieve superior classification performance compared to traditional logistic regression models (Dumitrescu et al., 2022).

Machine learning (ML) is increasingly being applied to forecasting challenges in economics and finance, emerging as a more effective and practical approach. Financial institutions, including banks and fintech companies, have been incorporating ML techniques into their operations (Li et al., 2024; Nazemi & Fabozzi, 2024; Dumitrescu et al., 2022; Kellner et al., 2022). This trend is also reflected in the literature on LGD and RR, where recent studies have compared traditional regression methodologies with newer ML techniques such as XGBoost, neural networks (NN), and random forests (RF).

Bellotti et al. conducted a study that explored the forecasting of recovery rates for NPLs by evaluating 20 models using regression and machine learning techniques. The models examined included seven linear models—ordinary least squares (OLS) regression, OLS regression with backward stepwise selection, two variations of ridge and lasso regression, and elastic net regression—as well as non-linear models such as multivariate adaptive regression splines (MARS), K-nearest neighbours (K-NN), model averaging neural networks, support vector machines (SVM), relevance vector machines, and Gaussian processes. Additionally, rule-based models were tested, including regression trees, conditional inference trees, bagged trees, random forests (RF), boosted trees, and cubist models. The research concluded that random forests, boosted trees, and cubist models achieved the best forecasting performance (Bellotti et al., 2021).

Li et al. provided a comprehensive summary of methodologies used for estimating LGD, categorizing them into three main groups: 1) statistical regression models, which are commonly employed, with Ordinary Least Squares (OLS) serving as a benchmark for LGD estimation (including OLS, fractional response regression, and beta regression); 2) machine learning and optimization algorithms, which are more powerful but often challenging to interpret (such as neural networks, support vector machines, and random forests); and 3) mixture and multi-stage models, which have been proposed to enhance LGD prediction (including inflated beta regression, zero-adjusted gamma, and two-stage models) (Li et al., 2021).

Xia et al. introduced a Heterogeneous Stacking Ensemble (HSE) approach for modelling LGD, demonstrating superior predictive performance compared to various other methodologies; however, it also presented increased complexity. In their research, HSE was evaluated against Ordinary Least Squares (OLS) regression, ridge regression, regression trees, multi-layer perceptrons (MLP), random forests (RF), Support Vector Regression (SVR), Gradient-Boosting Decision Trees (GBDT), XGBoost, LightGBM, CatBoost, and simple averaging (Xia et al., 2021).

The groundbreaking study by Nazemi et al. was the first to use deep learning to model collection rates and predict RR. These authors evaluated the effectiveness of various methodologies, including OLS regression, regression trees, support vector regression, deep learning, and boosting techniques, ultimately determining that deep learning outperformed all other approaches (Nazemi et al., 2022).

In the study conducted by Li et al., the authors concluded that tree-based non-parametric models significantly outperform parametric models, with XGBoost demonstrating the highest predictive ability. This conclusion was drawn from a comprehensive comparison of the results produced by four parametric models (OLS, RiR, LASSO, and Logit-Beta) and four tree-based non-parametric models (RF, AdaBoost, GBRT, and XGBoost) (Li et al., 2022).

In a recent study, it was found that classification trees and random forests achieved the highest accuracy compared with machine learning tree-based algorithms (such as decision trees, random forests, light gradient boosting, and XGBoost) and neural network algorithms (including backpropagation neural networks and extreme learning machines). Notably, XGBoost and extreme learning machine techniques exhibited the best performance (Li et al., 2024).

Malikkidou and Strohbach conducted a recent study that offers valuable insights for our research methodology. They developed an early warning system to predict distress for large European banks, comparing the performance of logistic regression with three machine learning methods: random forests, decision trees, and neural networks. Their findings indicate that, despite the traditional use of logistic regression models in early warning system design, machine learning techniques outperform traditional methods. Among these, the random forest model consistently demonstrates superior performance, while decision tree models closely follow, yielding comparable results across various performance metrics. Furthermore, their research highlights that the most influential variables in the early warning system are closely associated with bank profitability, including average interest expense on deposits and the asset-deposit spread, as well as solvency indicators such as equity to total liabilities and equity (Malikkidou & Strohbach, 2025).

Regarding the determinants of LGD and RR, there is no consensus in the literature on the most significant factors. Additionally, studies in this field often yield contradictory conclusions (Xu et al., 2024; Radivojević et al., 2019).

Returning to the academic literature on credit scoring, we identify five primary information sources integral to scoring systems, commonly referred to as the 5 Cs of credit: Character (the borrower's willingness to repay their debt), Capacity (the financial ability of the borrower to meet repayment obligations), Capital and collateral (the assets or equities of the borrower that can be liquidated for payment), and Conditions (which pertain to the overall economic environment or specific circumstances affecting the borrower or type of credit) (Rosenberg & Gleit, 1994).

In a similar vein to the initial 5 Cs of credit, our review of the literature on the determinants of LGD and RR reveals four key sources of information regarding these determinants: borrower characteristics, which encompass the principal traits of the debtor at the time credit was extended and how these have evolved, such as age, educational background, and sector of activity; debt or account characteristics, which involve all variables related to the debt, account, or contract, including debt amount, product type, and the presence of collateral; macroeconomic variables, such as GDP, unemployment rate, and interest rates; and repayment behaviour and C&R activities, which cover the actions taken by creditors to recover unpaid debts, including trace/no trace, litigation, and the number of late payments. We have compiled references related to these four primary sources in Table 2.

Table 2. Summary of data sources references for LGD and RR determinants

Data sources	Authors
Borrower characteristics	Xu et al., 2024; Li et al., 2024; Li et al., 2023A; Li et al., 2023B; Nazemi et al., 2022; Li et al., 2022; Calabrese & Zanin, 2022; Bellotti et al., 2021; Wang et al., 2020; Ye & Bellotti, 2019; Shoghi, 2019; Serrano-Cinca et al., 2015; Bijak & Thomas, 2015; Han & Jang, 2013; Bellotti & Crook, 2012; Hoechstoetter et al., 2012; Khieu et al., 2012; Höcht & Zagst, 2010; Chen et al., 2009; Chalupka & Kopecsni, 2008; Grunert & Weber, 2005
Debt or account characteristics	Xu et al., 2024; Li et al., 2024; Li et al., 2023A; Li et al., 2023B; Nazemi et al., 2022; Li et al., 2022; Calabrese & Zanin, 2022; Bellotti et al., 2021; Wang et al., 2020; Ye & Bellotti, 2019; Shoghi, 2019; Serrano-Cinca et al., 2015; Bijak & Thomas, 2015; Han & Jang, 2013; Bellotti & Crook, 2012; Hoechstoetter et al., 2012; Khieu et al., 2012; Höcht & Zagst, 2010; Chen et al., 2009; Chalupka & Kopecsni, 2008; Grunert & Weber, 2005
Macroeconomic variables	Xu et al., 2024; Li et al., 2023A; Li et al., 2023B; Nazemi et al., 2022; Li et al., 2022; Bellotti et al., 2021; Wang et al., 2020; Bijak & Thomas, 2015; Han & Jang, 2013; Bellotti & Crook, 2012; Khieu et al., 2012; Höcht & Zagst, 2010; Grunert & Weber, 2005
Repayment behaviour and C&R activities	Xu et al., 2024; Li et al., 2024; Li et al., 2023A; Li et al., 2023B; Nazemi et al., 2022; Li et al., 2022; Calabrese & Zanin, 2022; Bellotti et al., 2021; Wang et al., 2020; Ye & Bellotti, 2019; Shoghi, 2019; Serrano-Cinca et al., 2015; Bijak & Thomas, 2015; Han & Jang, 2013; Bellotti & Crook, 2012; Hoechstoetter et al., 2012; Khieu et al., 2012; Höcht & Zagst, 2010; Chen et al., 2009; Grunert & Weber, 2005

Source: Own elaboration

Basel II recommends that banks account for macroeconomic downturn conditions when predicting LGD and RR. Numerous studies have examined the correlation between these factors, yielding mixed results and no consensus. However, most findings suggest that incorporating macroeconomic variables enhances predictive accuracy.

Utilizing these macroeconomic variables helps capture the impact of business cycles on loan repayment and collection processes (Xia et al., 2021). Furthermore, economic uncertainty is considered a critical systematic determinant of RR distributions (Gambetti et al., 2019). Some studies have even concluded that the determinants of RR differ between "good" and "bad" times within the credit cycle, indicating that loan characteristics play a more significant role during prosperous periods, while specific firm characteristics and collateral determinants become more influential in challenging times (Wang et al., 2020).

The macroeconomic factors most frequently examined in studies include the unemployment rate, which typically exhibits a negative relationship with RR (Galow et al., 2024; Xia et al., 2021; Calabrese, 2012). In contrast, gross domestic product (GDP) generally demonstrates a positive relationship with RR (Galow et al., 2024; Radivojević et al., 2019; Friedman & Sandow, 2005). Interest rates usually have a negative correlation with RR (Galow et al., 2024; Xia et al., 2021). Stock market returns are generally positively related to RR (Galow et al., 2024; Bruche & González-Aguado, 2010), while investment growth also tends to show a positive association with RR (Bruche & González-Aguado, 2010). Household final consumption expenditure typically shows a negative relationship with RR (Radivojević et al., 2019). In contrast, the consumer confidence index shows mixed results, with either a negative or a positive correlation with RR (Xia et al., 2021). Table 3 summarizes the primary literature on the inclusion of these macroeconomic variables and their impact on model predictive power.

Table 3. Summary of macroeconomic variables used in the literature and their effect on models

Authors	Macroeconomic variables	Improves prediction flag
Wu et al., 2024	Consumer price index	Yes
Xu et al., 2024	Stock index, consumer price index, purchasing managers index, monetary aggregates growth	Yes
Galow et al., 2024	GDP, unemployment rate, stock returns, house prices, and long-term interest rates	Yes
Li et al., 2023A	Prime interest rates, producer price index, unemployment rate, and average real wage index	Yes
Li et al., 2023B	Lending interest rate, average wage growth, unemployment, producer price index, and industrial production index	Yes
Nazemi et al., 2022	Unemployment rate and excessive indebtedness rate	No
Kellner et al., 2022	20 variables for the USA loans and 18 for the Europe loans	Yes, but limited impact

Xia et al., 2021	3-month Treasury Bill rate, unemployment rate, average hourly earnings, and consumer confidence index	Yes
Bellotti et al., 2021	More than 400 macroeconomic, financial, structural, and housing market indicators from the European System of Central Banks (ESCB)	Yes
Li et al., 2021	S&P index	Yes
Wang et al., 2020	GDP, Industry median stock returns, and Credit spread	Yes
Radivojević et al., 2019	GDP, unemployment rate, inflation rate, household finale consumption expenditure, banks capital to assets ratio, and lending interest rate	Yes
Nazemi & Fabozzi, 2018	179 variables	Yes
Beck et al., 2017	GDP and unemployment rate	Yes, but limited impact
Han & Jang, 2013	GDP	Yes, but limited impact
Calabrese, 2012	GDP and unemployment rate	Yes
Bellotti & Crook, 2012	Bank interest rates, unemployment rate, and earnings growth (whole economy)	Yes
Khieu et al., 2012	GDP	Yes
Bruche & González-Aguado, 2010	GDP and credit cycle	Yes
Höcht & Zagst, 2010	GDP, TED spread, S&P index, Dow Jones Euro Stoxx index, growth rate in industrial production, volatility index from CBOE, VSTOXX and VDAX, treasury constant maturity rate, Euribor rate, consumer and price index, and S&P corporate default rate	Yes
Dermine & De Carvalho, 2006	GDP	No
Grunert & Weber, 2005	GDP, inflation rate, depreciations of banks quota, and interbank interest rate	Not confirmed
Friedman & Sandow, 2005	S&P-rated bonds default rate, S&P-rated bonds industry, and GDP	Yes
Schuermann, 2004	Business cycle moment: recession or expansion	Yes

Source: Own elaboration

The characteristics of borrowers and their debt or account attributes constitute broader categories of determinants that influence LGD and RR and carry significant weight in the statistical models used to predict them. This category encompasses various factors including whether the borrower is a natural or legal person, the number of days in arrears, instances of restructured loans, the debtor's age and income, residential status, marital status, educational background, address details, duration of relationship with the bank, type of financial product, years residing at the current address, employment status, credit bureau score, agency ratings, age of the defaulted account, loan amount, loan term, and the presence of multiple loans and/or creditors. Additionally, it considers attributes such as firm size, industry, age, legal structure, financial ratios, collateral or personal guarantees, type of collateral, debtor address region, borrower's indebtedness, EAD, Loan-to-Value (LTV) ratio, overall risk level, purpose of the loan, currency type, loan interest rate, customer outlook, country of residence or incorporation, instances of financial hardship, borrower affordability, and borrower viability. A summary of the most pertinent references found in the relevant literature regarding these variables is presented in Table 4.

Table 4. Summary of borrowers and debt variables in the literature

Data source	Variables	Authors
Borrowers' characteristics	Natural and legal person	Kriebel & Yam, 2020; Beck et al., 2017; ECB, 2017; Hoechstoetter et al., 2012
	Age of the debtor/firm age	Xu et al., 2024; Li et al., 2024, Nazemi et al., 2022; Li et al., 2022; Bellotti et al., 2021; Kriebel & Yam, 2020; Ye & Bellotti, 2019; Shoghi, 2019; Athreya et al., 2018; Thomas et al., 2012; Bellotti & Crook, 2012; Hoechstoetter et al., 2012; Zhang & Thomas, 2012; Bastos, 2010; Thomas et al., 2009; Chalupka & Kopecsni, 2008; Dermine & De Carvalho, 2006; Grunert & Weber, 2005
	Income and expenditure of the debtor	Xu et al., 2024; Li et al., 2023A; Li et al., 2023B; Calabrese & Zanin, 2022; Xia et al., 2021; Li et al., 2021; Shoghi, 2019; Athreya et al., 2018; Serrano-Cinca et al., 2015; Bellotti & Crook, 2012; Zhang & Thomas, 2012
	Address, postal code, region/province, and country of residence/incorporation	Xu et al., 2024; Li et al., 2024; Nazemi et al., 2022; Li et al., 2022; Calabrese & Zanin, 2022; Bellotti et al., 2021; Kriebel & Yam, 2020; Shoghi, 2019; ECB, 2017; Calabrese, 2012; Bellotti & Crook, 2012; Hoechstoetter et al., 2012; Höcht & Zagst, 2010; Chen et al., 2009; Grunert & Weber, 2005

	Years at current address	Bellotti & Crook, 2012; Zhang & Thomas, 2012; Matuszyk et al., 2010
	Residential status or asset presence	Xu et al., 2024; Li et al., 2023A; Li et al., 2023B; Nazemi et al., 2022; Li et al., 2022; Calabrese & Zanin, 2022; Serrano-Cinca et al., 2015; Thomas et al., 2012; Bellotti & Crook, 2012; Hoechstoetter et al., 2012; Zhang & Thomas, 2012; Thomas et al., 2009
	Gender	Xu et al., 2024; Li et al., 2024; Li et al., 2022; Bellotti et al., 2021; Kriebel & Yam, 2020; Ye & Bellotti, 2019; Hoechstoetter et al., 2012
	Marital status	Xu et al., 2024; Li et al., 2022; Bellotti et al., 2021; Ye & Bellotti, 2019; Zhang & Thomas, 2012
	Educational background	Xu et al., 2024; Li et al., 2022; Athreya et al., 2018
	Employment status, reference, experience, size or location	Xu et al., 2024; Li et al., 2023B; Calabrese & Zanin, 2022; Bellotti et al., 2021; Xia et al., 2021; Kriebel & Yam, 2020; Ye & Bellotti, 2019; Athreya et al., 2018; Serrano-Cinca et al., 2015; Bellotti & Crook, 2012; Zhang & Thomas, 2012
	Credit score, Credit bureau score, Agencies rating, Application score, PD, and internal scores, ratings and grades	Xu et al., 2024; Li et al., 2024; Li et al., 2023A; Li et al., 2023B; Nazemi et al., 2022; Bellotti et al., 2021; Xia et al., 2021; Li et al., 2021; Kriebel & Yam, 2020; Ye & Bellotti, 2019; Shoghi, 2019; Cheng & Cirillo, 2018; Tanoue et al., 2017; Serrano-Cinca et al., 2015; Han & Jang, 2013; Hoechstoetter & Nazemi, 2013; Thomas et al., 2012; Bellotti & Crook, 2012; Hoechstoetter et al., 2012; Khieu et al., 2012; Bastos, 2010; Matuszyk et al., 2010; Thomas et al., 2009; Chen et al., 2009; Dermine & De Carvalho, 2006; Bailey, 2002; Gupton, 2000
	Time or credit history with bank/creditor	Li et al., 2023A; Li et al., 2021; Shoghi, 2019; Bellotti & Crook, 2012; Zhang & Thomas, 2012; Bastos, 2010; Dermine & De Carvalho, 2006; Grunert & Weber, 2005; Bailey, 2002
	Credit history/inquiries	Xu et al., 2024; Li et al., 2023A; Li et al., 2023B; Calabrese & Zanin, 2022; Li et al., 2021; Serrano-Cinca et al., 2015; Hoechstoetter et al., 2012; Khieu et al., 2012; Bailey, 2002
	Level of risk and Customer outlook	ECB, 2017; Bailey, 2002
	Borrowers' indebtedness, affordability, or viability	Li et al., 2023A; ECB, 2017; Serrano-Cinca et al., 2015; Zhang & Thomas, 2012; Höcht & Zagst, 2010; Grunert & Weber, 2005

	Presence and number of investors	Xu et al., 2024; Xia et al., 2021
	Borrower group, lending group, or consumer family, and number of members	Li et al., 2022; Xia et al., 2021; Calabrese, 2012; Zhang & Thomas, 2012; Matuszyk et al., 2010
	Firm size, age, and status	Wang et al., 2020; Han & Jang, 2013; Khieu et al., 2012; Höcht & Zagst, 2010; Chen et al., 2009; Grunert & Weber, 2005
	Firm industry and industry characteristics	Xu et al., 2024; Jacobs, 2024; Li et al., 2022; Kellner et al., 2022; Hoechstoetter & Nazemi, 2013; Khieu et al., 2012; Qi & Zhao, 2011; Bastos, 2010; Bruche & González-Aguado, 2010; Höcht & Zagst, 2010; Chalupka & Kopecsni, 2008; Dermine & De Carvalho, 2006; Grunert & Weber, 2005; Schuermann, 2004; Gupton, 2000
	Firm legal form	Kellner et al., 2022; Kriebel & Yam, 2020; Höcht & Zagst, 2010; Grunert & Weber, 2005
	Firm financial ratios	Kriebel & Yam, 2020; Wang et al., 2020; Han & Jang, 2013; Khieu et al., 2012; Grunert & Weber, 2005; Altman, 1968
Debt or account characteristics	Arrears days past due and default date	Li et al., 2024; Nazemi et al., 2022; Bellotti et al., 2021; Xia et al., 2021; Kriebel & Yam, 2020; Wang et al., 2020; ECB, 2017; Bijak & Thomas, 2015; Han & Jang, 2013; Hoechstoetter & Nazemi, 2013; Thomas et al., 2012; Bellotti & Crook, 2012; Zhang & Thomas, 2012; Matuszyk et al., 2010; Höcht & Zagst, 2010; Thomas et al., 2009; Chalupka & Kopecsni, 2008; Bailey, 2002
	Type of product	Nazemi et al., 2022; Kellner et al., 2022; Bellotti et al., 2021; Wang et al., 2020; Ye & Bellotti, 2019; Shoghi, 2019; Hoechstoetter & Nazemi, 2013; Zhang & Thomas, 2012; Khieu et al., 2012; Qi & Zhao, 2011; Höcht & Zagst, 2010; Chen et al., 2009; Chalupka & Kopecsni, 2008; Bailey, 2002; Gupton, 2000
	Product/loan amount or EAD	Xu et al., 2024; Li et al., 2024; Jacobs, 2024; Li et al., 2023A; Li et al., 2023B; Nazemi et al., 2022; Li et al., 2022; Calabrese & Zanin, 2022; Bellotti et al., 2021; Xia et al., 2021; Li et al., 2021; Kriebel & Yam, 2020; Wang et al., 2020; Ye & Bellotti, 2019; Shoghi, 2019; Cheng & Cirillo, 2018; Beck et al., 2017; Tanoue et al., 2017; ECB, 2017; Serrano-Cinca et al., 2015; Bijak & Thomas, 2015; Han & Jang, 2013; Hoechstoetter & Nazemi, 2013; Thomas et al., 2012;

		Calabrese, 2012; Bellotti & Crook, 2012; Hoechstoetter et al., 2012; Khieu et al., 2012; Bastos, 2010; Matuszyk et al., 2010; Höcht & Zagst, 2010; Thomas et al., 2009; Chalupka & Kopecsni, 2008; Dermine & De Carvalho, 2006; Grunert & Weber, 2005; Bailey, 2002
	Collateral or personal guarantee, type and location of collateral, or product/collateral seniority	Jacobs, 2024; Li et al., 2022; Kellner et al., 2022; Wang et al., 2020; Tanoue et al., 2017; ECB, 2017; Hoechstoetter & Nazemi, 2013, Calabrese, 2012; Khieu et al., 2012; Qi & Zhao, 2011; Bastos, 2010; Höcht & Zagst, 2010; Chen et al., 2009; Chalupka & Kopecsni, 2008; Dermine & De Carvalho, 2006; Grunert & Weber, 2005; Friedman & Sandow, 2005; Schuermann, 2004; Gupton, 2000
	Presence or number of multiple loans, products, and/or multiple creditors	Xu et al., 2024; Jacobs, 2024: Li et al., 2023A; Li et al., 2023B; Nazemi et al., 2022; Li et al., 2022; Calabrese & Zanin, 2022; Li et al., 2021; Kriebel & Yam, 2020; Ye & Bellotti, 2019; Serrano-Cinca et al., 2015; Bellotti & Crook, 2012; Zhang & Thomas, 2012; Höcht & Zagst, 2010; Chalupka & Kopecsni, 2008; Grunert & Weber, 2005; Bailey, 2002; Gupton, 2000
	Product age	Bellotti et al., 2021; Beck et al., 2017; Hoechstoetter et al., 2012; Chalupka & Kopecsni, 2008
	Purpose of credit/loan	Xu et al., 2024; Li et al., 2023B; Calabrese & Zanin, 2022; Xia et al., 2021; ECB, 2017; Serrano-Cinca et al., 2015; Zhang & Thomas, 2012; Chen et al., 2009
	Product / loan interest rate	Xu et al., 2024; Li et al., 2024; Li et al., 2023A; Li et al., 2022; Calabrese & Zanin, 2022; Bellotti et al., 2021; Ye & Bellotti, 2019; Athreya et al., 2018; ECB, 2017; Bastos, 2010; Chalupka & Kopecsni, 2008; Dermine & De Carvalho, 2006
	Product / loan term	Xu et al., 2024; Calabrese & Zanin, 2022; Xia et al., 2021; Bijak & Thomas, 2015; Thomas et al., 2012; Zhang & Thomas, 2012; Thomas et al., 2009; Grunert & Weber, 2005
	Product / loan fees	Bellotti et al., 2021; Ye & Bellotti, 2019; Athreya et al., 2018
	Credit limit and limit utilization	Bellotti et al., 2021; Li et al., 2021; Ye & Bellotti, 2019; Höcht & Zagst, 2010; Bailey, 2002
	Revolving utilization and balance	Li et al., 2023A; Li et al., 2023B; Calabrese & Zanin, 2022; Xia et al., 2021; Li et al.,

		2021; Serrano-Cinca et al., 2015; Bailey, 2002
	Insurance presence	Athreya et al., 2018; Zhang & Thomas, 2012; Bailey, 2002
	LTV	Cheng & Cirillo, 2018; ECB, 2017; Hoechstoetter & Nazemi, 2013; Chalupka & Kopecsni, 2008
	Re-structured, bankruptcy, hardship or insolvent cases	Li et al., 2023A; Li et al., 2023B; Kriebel & Yam, 2020; Wang et al., 2020; Beck et al., 2017; ECB, 2017; Khieu et al., 2012; Chen et al., 2009; Bailey, 2002
	Reason for default	Höcht & Zagst, 2010
	Currency	ECB, 2017
	Client portfolio	Nazemi et al., 2022

Source: Own elaboration

In recent years, there has been a growing emphasis on repayment behaviour and actions taken in the C&R process, particularly on their contributions to research on RR and LGD. Information related to the C&R process has been shown to enhance the performance of LGD and RR models significantly. This type of information encompasses various data points, including the duration of centralized management, the involvement of external agencies, the initiation of litigation, whether or not there is traceability of debtors, the duration and costs associated with the workout process, credit bureau information, communication records (calls and contacts), the reasons for default, and repayment behaviour statistics (such as the number of late payments, average late payment duration, payment discrepancies, and cumulative differences in payments). Notably, researchers have placed special emphasis on the importance of post-default information for further enhancing the predictive capabilities of these models. Table 5 summarizes the primary references to these variables in the existing literature.

Table 5. Summary of repayment behaviour and C&R activities variables in the literature

Variables	Authors
Days of centralized or outsourced management	Li et al., 2022; Kriebel & Yam, 2020; Shoghi, 2019; Höcht & Zagst, 2010
Presence or experience with external agencies	Li et al., 2022; Beck et al., 2017
The start of litigation process and its status	Li et al., 2022; Bailey, 2002

Trace or no trace of debtors	Matuszyk et al., 2010; Bailey, 2002
Duration and costs of the workout or resolution process	Jacobs, 2024; Tanoue et al., 2017; Chalupka & Kopecsni, 2008; Grunert & Weber, 2005; Bailey, 2002; Gupton, 2000
Credit bureau data	Xia et al., 2021; Ye & Bellotti, 2019; Bailey, 2002
Calls, contacts and visits (number, results and timings)	Bellotti et al., 2021; Ye & Bellotti, 2019; Shoghi, 2019; Bailey, 2002
Contact information (mobile phone, fixed-line phone, number of phones…)	Nazemi et al., 2022; Kriebel & Yam, 2020; Shoghi, 2019; Thomas et al., 2012; Thomas et al., 2009; Bailey, 2002
Type of bankruptcy process	Wang et al., 2020; Han & Jang, 2013
Repayment behaviour data (number of late payments, average number of late payments, differences of payments and cumulative differences, debt settlements)	Xu et al., 2024; Li et al., 2024; Li et al., 2023A; Nazemi et al., 2022; Calabrese & Zanin, 2022; Bellotti et al., 2021; Kriebel & Yam, 2020; Wang et al., 2020: Ye & Bellotti, 2019; Shoghi, 2019; Cheng & Cirillo, 2018; Beck et al., 2017; ECB, 2017; Thomas et al., 2016; Serrano-Cinca et al., 2015; Bijak & Thomas, 2015; Han & Jang, 2013; Bailey, 2002
Time between default and sell-off to third-party or date of sell-off	Nazemi et al., 2022; Hoechstoetter et al., 2012

Source: Own elaboration

(3) The Segmentation Relevance

Segmentation has been recognized in the literature as having significant relevance for C&R and LGD in two primary ways: optimizing C&R outcomes and enhancing modelling performance.

In general, customer segmentation involves grouping customers with similar attributes, enabling the development of specific, differentiated strategies tailored to each group or segment (Alpaydin, 2021). When applied to C&R activities, segmentation improves debt collection performance by facilitating the grouping of borrowers with comparable characteristics, thereby enabling the implementation of similar treatments and customized processes for each segment. In some instances, specialized expert teams are even assigned to these groups (ECB, 2017; Chin & Kotak, 2006; Coleman, 2004).

Sophisticated segmentation is recognized as one of the most effective techniques for enhancing collection results. Industry best practices acknowledge that not all customers are alike; they identify various types of debtors and address them accordingly.

Behavioural and collection scoring methods predict the likelihood that a delinquent account will repay in the future. Profiles or segments are established based on these scores, and distinct actions are taken for groups of accounts that exhibit significantly different levels of risk (Bailey, 2002).

Behavioural segmentation of debt, informed by psychological insights and advanced analytics tailored to each stage of the collections process, has resulted in improvements of 20-30% in the amounts collected and reductions in write-offs for lending institutions that employ these techniques (Baer, 2018).

Deville's study on collection success highlights the importance of affect management in the relationship between debtors and collectors. He identifies two fundamental components of consumer collections strategies: first, customer base segmentation, which allows for the application of tailored collections strategies to distinct groups of debtors (determining how and when to make contact); and second, the implementation of champion vs. challenger testing in collections and recovery management activities (Deville, 2014).

Significant resources and costs are dedicated to debt collection. Enhancing the identification of clients or accounts that are more likely to repay their debts, along with a prioritization management process, leads to a reduction in collections costs and improved efficiency of collections and servicing efforts (Malinconico & Virguti, 2024; Kriebel & Yam, 2020; Drozd & Serrano-Padial, 2017; Drozd & Serrano-Padial, 2014).

Efficiency in collections hinges on aligning the right actions with each debtor, achieving optimal cost when the marginal savings from write-offs equal the incremental costs of taking action (Bailey, 2002). Not all accounts should be treated uniformly; they are ranked and categorized by risk level: high-risk accounts are contacted more frequently, low-risk accounts receive minimal attention, and all others undergo a standardized process. This segmentation strategy helps avoid the costs of collecting from borrowers who are unable to repay, focuses collection efforts on delinquent borrowers who are more likely to pay, and enables aggressive collection strategies for higher-risk accounts. The market has reported productivity gains of 20-40% by applying segmentation and prioritization in collections (Malinconico & Virguti, 2024; Drozd & Serrano-Padial, 2017; Thomas et al., 2016; Lawrence, 1992).

In addition to enhancing performance and efficiency, applying segmentation to C&R activities reduces customer complaints (Lawrence, 1992).

As previously mentioned, segmentation in the literature is deemed crucial for enhancing the performance of both RR and LGD modelling (Carleo & Rocci, 2024; Thomas & Bijak, 2015). A more accurate estimate of the recovery performance of an NPL portfolio can be obtained by dividing the portfolio into several homogeneous subsets, or clusters, enabling the calculation of a recovery curve for each cluster (Carleo & Rocci, 2024). Furthermore, LGD models can be enhanced by segmenting the population and developing distinct models for each segment, as various subgroups may exhibit different characteristics and distributions (Li et al., 2024; Zhang & Thomas, 2012). These methodological improvements in screening for defaulting borrowers are expected not only to influence C&R results positively but also to contribute to more effective pricing and availability in revolving credit (Fedaseyeu & Hunt, 2015).

Additionally, Shoghi has developed a machine learning model—specifically, a gradient-boosted decision tree—to optimize the collection process for a Debt Collection Agency (DCA) by prioritizing which debtors to call to maximize long-term recovery of delinquent debt. This model calculates the expected value of each debtor, factoring in both the likelihood of repayment and the size of the debt, ultimately computing the marginal value of a phone call for each debtor. Consequently, debtors with the highest marginal value per call are prioritized. The implementation of this model resulted in a notable performance increase, with a 14% rise in collected debt and a 22% reduction in calling effort, and it also proved helpful in predicting collection performance and resource requirements (Shoghi, 2019).

(4) The COVID-19 effect

In the literature, there is a growing body of studies analysing the effects of COVID-19 on society and various economic sectors; however, there appears to be a lack of research specifically addressing its impact on the C&R business and on debt segmentation.

COVID-19 first emerged in December 2019, prompting the World Health Organization (WHO) to declare it a pandemic and an international public health emergency the following month. By March 2020, the epidemic had spread globally. As of April 13, 2024, there have been over 700 million confirmed cases and 7 million deaths related to the virus (data sourced from https://www.worldometers.info/coronavirus/) (Lin & Lu, 2024). The pandemic and the measures implemented to curb its spread triggered a worldwide economic recession, marking a transition from a prolonged period of benign credit conditions to a more stressed environment. However, the effects varied across countries and sectors. In Spain, the impact was particularly significant, as it was one of the first countries, alongside China and Italy, to experience outbreaks. Confinement measures were enacted starting March 15, 2020, and lasted until June 21, 2020, leading to a severe disruption of economic activity in the Spanish market (Lin & Lu, 2024; Ramos Gonzalez et al., 2023; Altman, 2021).

The limited literature examining the impact of COVID-19 on the C&R industry includes research highlighting public and legal measures implemented by regulators and public institutions in response to the pandemic, specifically regarding debt collection practices. Brakes analyse the public actions taken regarding debt collection in Idaho, USA, identifies notable gaps in these measures, and offers recommendations for government, legal, and judicial bodies for future initiatives (Brakes, 2020). In another study, O'Brien evaluates the legal framework governing the debt collection industry in Australia following COVID-19 and presents recommendations to safeguard consumers from potentially harmful debt collection practices (O'Brien et al., 2021).

A recent report from the European Banking Authority (Malikkidou & Strohbach, 2025) indicated that the overall impact of COVID-19 on the banking sector was less severe than initially anticipated, mainly due to the implementation of public support measures for individuals and businesses. These measures varied significantly across countries and included lines of credit, loan guarantees, moratoriums, job protection schemes, and tax deferrals (Cowling et al., 2022). However, recent studies examining the determinants of loan defaults and credit losses have concluded that the impact of COVID-19 on these areas was indeed significant.

In their research, Ramos Gonzalez et al. used a machine learning algorithm to forecast the expected loss for defaulted mortgages at the three largest credit institutions in Spain, accounting for the impact of COVID-19 (Ramos Gonzalez et al., 2023). The primary conclusion of this study is that the pandemic crisis has significantly negatively affected the analysed portfolios of defaulted mortgages. However, the impact varied notably across the three entities under investigation. The authors also emphasized that the economic repercussions of the COVID-19 crisis in Spain were akin to those experienced during the global financial crisis that began in 2008.

Meanwhile, Xu et al. explored the determinants of loan defaults in peer-to-peer (P2P) lending both before and during the COVID-19 pandemic using machine learning methods (Xu et al., 2024). This study drew on four data sources: loan characteristics (such as interest rate, loan amount, loan term, loan purpose, remaining period, repayment source, and bidders), credit transaction history (including credit rating, credit score, application frequency, successful application counts, success rates, total borrowed amount, principal and interest, and repayment counts), personal demographic information (covering age, gender, marital status, education, income, work experience, residence region, industry, asset ownership, debt levels, and property size and location), and macroeconomic indicators (like the China stock index 300, consumer price index, M0, M1, M2 growth, and purchasing managers index). The findings revealed a significant impact of COVID-19 on both the most relevant predictors of loan defaults and the influence some predictors had on loan default behaviour (either promoting or reducing it). The top five predictors of loan default prior to COVID-19, ranked by relevance, were principal and interest, purchasing managers index, repayment times, loan amount, and total borrowed amount. In contrast, during the pandemic, the most significant variables shifted to principal and interest, credit score, remaining period, repayment times, and loan amount.

Research Gap and Contribution to the Literature

While the existing literature provides valuable insights into C&R activities, methodologies, and the determinants of RR and LGD, we have not identified any studies that directly address the specific focus of our research.

To the best of our knowledge, this study represents the first exploration of the most effective segmentation variables and the impact of an event such as COVID-19 (and all the linked measures to mitigate its effect) on debt segmentation for collections and recoveries of unsecured written-off portfolios acquired by investment firms from banks and other financial institutions in the Spanish market.

This study contributes to the research area of the more relevant determinants for RR modelling estimations, as our models predict the probability of an account to make a payment in the following 30 days, directly linked to the RR, as explained in Section 3.3 of this document.

The results of our research align with recent developments in the literature regarding the relevance of the four sources of information (borrower characteristics, debt characteristics, macroeconomic variables, and repayment behaviour and C&R activities), as well as the increasing importance of the C&R activities variables. But our research goes further, emphasizing the relevance of this category of variables.

On the contrary, our results are completely distinct from the more traditional literature, which only notes the importance of borrower and debt characteristics and does not include the C&R activities variables.

It is relevant to highlight that our research uses data that is typically inaccessible to academic studies because it is not publicly available. The dataset is comprehensive, encompassing information from approximately 900,000 accounts and tracking their payment behaviour over six years, from 2018 to 2023. This database draws upon the four primary sources identified in the literature: borrower characteristics, debt characteristics, macroeconomic variables, and repayment behaviour and C&R activities.

SECTION 3.- DATA FOR THE RESEARCH

3.1.- DATA AVAILABLE

3.2.- DATA CLEANING AND NORMALIZATION

3.3.- VARIABLE TO BE EXPLAINED

3.4.- DATA SETS AND EXPLANATORY VARIABLES

3.1.- DATA AVAILABLE

For this research, we were granted access to data on portfolios acquired in the Spanish market by the Investment Firm and its affiliated entities (hereinafter, the "Investment Firm"). This Investment Firm is part of a Global Investment Firm Group that maintains a prominent presence and reputation in the market. A non-disclosure agreement has been established with the Investment Firm to govern the access and management of this data.

Founded in the 1990s, this Investment Firm has been a pioneer in alternative investments, boasting a global footprint. It holds investments totalling several billion US dollars in assets across various sectors, including complementary credit, private equity, and real estate platforms. Notably, it ranks among the world's leading investors in NPL portfolios.

In the Spanish market, the Investment Firm has been highly active over the past two decades, engaging extensively in the acquisition of NPL and real estate portfolios and enhancing its servicing capabilities. Its primary service offerings in Spain include amicable debt management, judicial debt management, acquisition of written-off debt portfolios, advisory services related to the acquisition of written-off debt portfolios, insolvency management, mortgage execution processes, and general consulting services (Informa, 2023).

The business model of this Investment Firm aligns with Malinconico & Virguti's classification as a hybrid model of debt acquisition, recovery, and third-party servicing, synthesizing elements of both debt acquisition and third-party servicing (Malinconico & Virguti, 2024).

The research incorporates data from approximately 900,000 accounts across 35 portfolios acquired by the Investment Firm in the Spanish market between 2015 and mid-2022. These accounts are categorized into two primary groups due to their distinct behaviours (Dumitrescu et al., 2022; ECB, 2017): Individual account holders, which comprise over

700,000 accounts and represent €3.3 billion in debt; and Small and medium-sized enterprises (SMEs), which account for more than 200,000 files and €10.9 billion in debt.

The database assembled for this research includes five key sources of information, akin to those found in the academic literature (Xu et al., 2024; Li et al., 2024; Li et al., 2023A; Khieu et al., 2012). The Investment Firm provided data on accounts, debtors, payment behaviours, and collection activities. At the same time, macroeconomic factors were sourced from the public information website of the Instituto Nacional de Estadística (INE, Spanish National Statistics Institute, www.ine.es). Detailed information about these data sources is provided in Table 6.

Table 6. Data fields included in the database

Data category	Data fields
Accounts characteristics at the portfolios acquisition	• Account ID unique number • Individuals or SMEs segment flag • Name of the acquired portfolio • Name of the credit originator • Debt type of product • Detail of the original debt: total amount, principal amount, interests amount and others amount • Detail of the actual debt: total amount, principal amount, interests amount and others amount • Debt delinquency date • Debt cancellation date • Debt upload by the investor date
Debtor characteristics	• Debtor ID unique number • Type of participant in the debt • Date of birth • Type of ID document • Number of mobile phone numbers • Number of fixed-line phone numbers • Postal code
Payment behaviour in the 2018-2023 period of time	• Payment unique ID number • Payment amount • Payment date • Method of payment • Management phase in which the payment was made
Collection activities delivered by the Investment Firm	• SMS sending unique ID number • SMS sending date • Communications model • Calls unique ID number • Number of calls by month

in order to recover the debt	• Number of useful contacts by month • Number of payments promises by month • Judicial information: Judicial procedure unique ID number, Judicial procedure status, Judicial claim date, Judicial "auto" number and Judicial procedure discharge date
Macroeconomic factors	• General Consumer Price Index (on annual variations and on monthly variations) • Underlying Consumer Price Index (on annual variations and on monthly variations) • Consumer Confidence Index (ICC, actual situation and expectations) • GDP Market Prices (annual and quarterly variations) • GDP Full-Time Equivalent Employment (annual and quarterly variations) • Unemployment rate (on annual variation) • Activity rate (on annual variation)

Source: Own elaboration

The information was sourced from 13 data files provided by the Investment Firm and 11 data files obtained from the INE website for the macroeconomic factors. Appendix 1 includes a description detailing the fields of these files, along with an example extract from each data set collected for the research.

3.2.- DATA CLEANING AND NORMALIZATION

Once the various files were received and analysed, we worked to construct the datasets for modelling. Each field of the data files was carefully reviewed, and several interactions were held with the Investment Firm to clarify uncertainties and ensure a comprehensive understanding of the data. Based on the significant differences in payment behaviour and modelling, two distinct datasets were created: one for Individuals and another for SMEs (Dumitrescu et al., 2022; ECB, 2017).

To build the data sets for the modelling, a process of data cleaning and normalization was performed, with the following main steps:

1. Join all the accounts files available, with a review and adjustment of fields to make the joint possible.
2. Group the different names of the portfolios acquired to properly unify and treat them, as in several cases, the same portfolio acquired had been named in slightly different ways. The Investment Firm files were shared twice, so this grouping exercise was done twice as well.
3. The names of the verticals of clients were transformed in order to make them more self-explanatory. This transformation is shown in the following table.

Table 7. Vertical field transformation

Original vertical name	Vertical renamed
I	Individual
E	SME

Source: Own elaboration

4. Several portfolios were eliminated from the data set due to the following reasons: some portfolios had a small size, some portfolios were of secured debt products (and this research is focused on unsecured debt products), by recommendation of the Investment Firm (due to certain portfolio characteristics which made them of low

representativeness and irrelevant) or for being portfolios with a relevant lack of data associated.

5. Information on the portfolio date of registration was added.
6. The fields were established to delineate distinct ranges, or categories, associated with the outstanding amounts. The outstanding monthly amount was calculated from the original amount and the cumulative movements through the month of the data point. The data was split into eight ranges of the outstanding amount: 0-100€, 100-500€, 500-1,000€, 1,000-3,000€, 3,000-10,000€, 10,000-100,000€, 100,000-1,000,000€, and >1,000,000€
7. Unify the different types of debt creditors and group them to have representative categories.
8. Unify the different types of debt products and group them into representative categories.
9. Information on the "Personas" (persons) files was added.
10. Date of birth fields were adjusted to leave in the data set only the minimum ("fecha_min") and maximum ("fecha_max") dates available.
11. Group and delineate specific fields to facilitate the establishment of ranges for the phone numbers variables.
12. Group and delineate specific fields to facilitate the establishment of ranges for the postal code variables.
13. Group and delineate specific fields to facilitate the establishment of ranges for the region ("Comunidades Autónomas") variables.
14. Fields were established to delineate various ranges concerning the number of individuals..
15. Fields were established with the nationality of persons.
16. Available judicial information was included in the data sets.

17. Portfolio of telecommunications ("TELCO") was eliminated. This portfolio was the only one of this type, and its debt type was not the focus of this research.
18. Available SMS information was included in the data sets.
19. Available call information was included in the data sets.
20. Based on the activity data, several fields were established to show the activity in the last 30, 60, and 90 days, and historical information.
21. Available payment information was included in the data sets.
22. Relevant fields about the payment information were calculated for each of the time periods to analyse (30 days, 60 days, 90 days, and 180 days).
23. The variable objective to predict fields were established. These fields were used as objective variables for modelling purposes.
24. Fields with the flag of payers in the last periods of time were established for each account (Three fields: payer in the last 30 days, payer in the last 90 days, and payer ever).
25. Clean and adjustment of fields and values on the fields to control the quality of the information: null values, fields with very low representativeness values…
    a. All the input variables were renamed to include the prefix "IN" and a correlative number, in order to facilitate their identification and management.
    b. All the output variables were renamed to include the prefix "OUT" and a correlative number, in order to facilitate their identification and management.
    c. The collection activities variables were transformed into dichotomous variables: 0 for a negative result and 1 in the case of a positive result.
    d. The payments output variables were transformed into dichotomous variables: 0 for non-payments and 1 for payments.

26. For each of the variables, univariate and bivariate calculations and analysis were made. Also, for each of them, the Information Value (IV) and the Weight of Evidence (WOE) were calculated.
27. A different data set of the two types of clients available is established: Individuals and SMEs.
28. The final variables for each of the data sets were the eighty-one included in the following table:

Table 8. List of variables in the data sets

Variable	Description
1. "IUC"	Account identifier
2. "NUMERO"	Account identifier
3. "SERIE"	Account identifier
4. "Fecha"	Date: month and year
5. "IN01_Cartera_G"	Portfolio
6. "IN02_Vertical"	Vertical segment
7. "IN03_Originador"	Creditor
8. "IN04_Tipo_Producto"	Product type
9. "IN05_Saldo_INI_G"	Initial total debt amount
10. "IN06_Saldo_INI_INT_G"	Initial interest debt amount
11. "IN07_Saldo_INI_NOM_G"	Initial principal debt amount
12. "IN08_Saldo_INI_OTH_G"	Initial other debt amount
13. "IN09_Mes_Impago"	Month and year of the original debt
14. "IN10_Anio_Impago"	Year of the original debt
15. "IN11_Originador_G"	Creditor grouped
16. "IN12_Tipo_Producto_G"	Product type grouped
17. "IN13_Nacimiento_Min"	Oldest date of birth or date of company creation of participants
18. "IN14_Nacimiento_Max"	Youngest date of birth or date of company creation of participants
19. "IN15_NUM_TEL_FIJO_Sum"	Number of fixed line phone numbers
20. "IN16_NUM_TEL_MOVIL_Sum"	Number of mobile phone numbers
21. "IN17_NE_EXT"	Foreign nationality flag
22. "IN18_NE_NAC"	National nationality flag
23. "IN19_NE_OTH"	Other nationality flag
24. "IN20_Provincia"	Province
25. "IN21_TP_TIT"	Holder or active company flag
26. "IN22_TP_AVA"	Guarantor flag
27. "IN23_TP_OTH"	Other participant or relationship flag
28. "IN24_NumPersonas"	Number of participants
29. "IN25_TotalTelefonos"	Number of phone numbers
30. "IN26_CCAA"	Region
31. "IN27_Judicial"	Judicial flag
32. "IN28_Saldo_InicioCorte_G"	Monthly debt amount grouped

33. “IN29_SMS_L30D” 34. “IN30_N_LLAMADAS_L30D” 35. “IN31_RPC_L30D” 36. “IN32_PTP_L30D” 37. “IN33_SMS_L90D” 38. “IN34_N_LLAMADAS_L90D” 39. “IN35_RPC_L90D” 40. “IN36_PTP_L90D” 41. “IN37_SMS_Ever” 42. “IN38_N_LLAMADAS_Ever” 43. “IN39_RPC_Ever” 44. “IN40_PTP_Ever” 45. “IN41_Pago_L30D” 46. “IN42_Pago_L90D” 47. “IN43_Pago_Ever” 48. “IN44_IPC_IndGeneral_VarAnual” 49. “IN45_IPC_IndGeneral_VarAnual_TIL10” 50. “IN46_IPC_IndGeneral_VarMensual” 51. “IN47_IPC_IndGeneral_VarMensual_TIL10” 52. “IN48_IPC_IndSubyacente_VarAnual” 53. “IN49_IPC_IndSubyacente_VarAnual_TIL10” 54. “IN50_IPC_IndSubyacente_VarMensual” 55. “IN51_IPC_IndSubyacente_VarMensual_TIL10” 56. “IN52_ICC” 57. “IN53_Ind_Sit_Actual” 58. “IN54_Ind_Expectativas” 59. “IN55_ICC_TIL10” 60. “IN56_Ind_Sit_Actual_TIL10” 61. “IN57_Ind_Expectativas_TIL10” 62. “IN58_Paro_TasaActividad_VarAnual” 63. “IN59_Paro_TasaActividad_VarAnual_TIL10” 64. “IN60_Paro_TasaParo_VarAnual” 65. “IN61_Paro_TasaParo_VarAnual_TIL10” 66. “IN62_PIB_EmpleoEquivTiempoCompleto_VarAnual” 67. “IN63_PIB_EmpleoEquivTiempoCompleto_VarAnual_TIL10”	SMS sent in the last 30 days flag Calls in the last 30 days flag Right-party contact in the last 30 days flag Promise to pay in the last 30 days flag SMS sent in the last 90 days flag Calls in the last 90 days flag Right-party contact in the last 90 days flag Promise to pay in the last 90 days flag SMS sent ever flag Calls ever flag Right-party contact ever flag Promise to pay ever flag Payment in the last 30 days flag Payment in the last 90 days flag Payment ever flag General consumer price index on annual variation General consumer price index on annual variation grouped in deciles General consumer price index on monthly variation General consumer price index on monthly variation grouped in deciles Underlying consumer price index on annual variation Underlying consumer price index on annual variation grouped in deciles Underlying consumer price index on monthly variation Underlying consumer price index on monthly variation grouped in deciles Consumer confidence index Consumer confidence index actual situation Consumer confidence index expectations Consumer confidence index grouped in deciles Consumer confidence index actual situation grouped in deciles Consumer confidence index expectations grouped in deciles Activity rate on annual variation Activity rate on annual variation grouped in deciles Unemployment rate on annual variation Unemployment rate on annual variation grouped in deciles

68. “IN64_PIB_EmpleoEquivTiempoCompleto_VarTrim” 69. “IN65_PIB_EmpleoEquivTiempoCompleto_VarTrim_TIL10” 70. “IN66_PIB_PreciosMercado_VarAnual” 71. “IN67_PIB_PreciosMercado_VarAnual_TIL10” 72. “IN68_PIB_PreciosMercado_VarTrim” 73. “IN69_PIB_PreciosMercado_VarTrim_TIL10” 74. “OUT01_Cob_60D” 75. “OUT02_Cob_90D” 76. “OUT03_Cob_180D” 77. “OUT04_Pago_60D” 78. “OUT05_Pago_90D” 79. “OUT06_Pago_180D” 80. “OUT07_Cobros_30D” 81. “OUT08_Pago_30D”	GDP full-time equivalent employment on annual variation GDP full-time equivalent employment on annual variation grouped in deciles GDP full-time equivalent employment on quarterly variation GDP full-time equivalent employment on quarterly variation grouped in deciles GDP market prices on annual variation GDP market prices on annual variation grouped in deciles GDP market prices on quarterly variation GDP market prices on quarterly variation grouped in deciles Amount paid in the following 60 days Amount paid in the following 90 days Amount paid in the following 180 days Payment in the following 60 days flag Payment in the following 90 days flag Payment in the following 180 days flag Amount paid in the following 30 days Payment in the following 30 days flag

Source: Own elaboration

3.3.- VARIABLE TO BE EXPLAINED

The variable analysed by the models is the probability that an account will make a payment within the next 30 days, calculated monthly and directly linked to the recovery rate (RR). RR is defined as the proportion of the amount outstanding that is successfully collected or recovered by the entities, after deducting administrative fees incurred during the collection process (Galow et al., 2024; Hoechstoetter & Nazemi, 2013; Repullo et al., 2010; BCBS, 2005).

The RR is used to estimate the final amount collected by multiplying it by the amount owed. In the case of our probability that an account will make a payment within the next 30 days, to estimate the final amount collected, we multiply the estimated probability by the ratio of the amount paid to the total amount owed. If the debt is paid in full, RR and our probability yield the same value; and if the debt is paid partially, our probability has to consider the ratio of the amount paid to the total amount owed when converting to the RR.

A review of the academic literature on RR estimations indicates that predicting RRs yields more accurate models than estimating recovery amounts directly. Furthermore, better predictions of the recovery amount can be achieved by first estimating the RR and then multiplying it by the default amount (Zhang & Tomas, 2012). This approach is also regarded as best practice among industry practitioners.

To define the variable to be analysed, up to eight potential options were studied, taking into account the prediction timeframe (payments in 30, 60, 90, or 180 days) and whether to estimate the exact recovery amount or predict the likelihood of an account making a payment. Ultimately, the decision was made to focus on whether an account will make a payment within the next 30 days, based on three primary factors.

First, and most relevant reason: the data utilized for this research captures monthly snapshots of portfolio behaviour, making it the most granular option available, and the models are executed monthly. So, timeframes longer than 30 days would be overridden by these monthly updates, increasing errors in the model, reducing its predictive power (the Gini coefficient would decrease), and overlapping with prioritisation of accounts to manage. Second, these models support portfolio management by estimating the probability of collection and the resources required (such as collector agents, calls, SMS, letters, etc.) based on established management activities and strategies. This process is typically monitored and reviewed monthly by the entities responsible for collection efforts. Lastly, in written-off portfolios, the recovery rate tends to be relatively low, and management efforts focus on "activating" debtors to encourage them to become payers, regardless of whether payments are partial or full.

3.4.- DATA SETS AND EXPLANATORY VARIABLES

The data sets for Individuals and SMEs include both input (explanatory) variables and output (to be explained) variables for the period from 2018 to 2023. This information pertains to the following three groups:

1. Input Data (Explanatory Variables) - information that varies for each account, including identifiers for accounts and individuals, characteristics of debtors and their debts, collection activities conducted by the Investment Firm, and historical payment behaviour associated with each account; and macroeconomic factors, which are consistent across accounts.
2. Output Data – concerning payment patterns over the next 30, 60, 90, and 180 days.
3. Dates – covering the period from 2018 to 2023, provided every month.

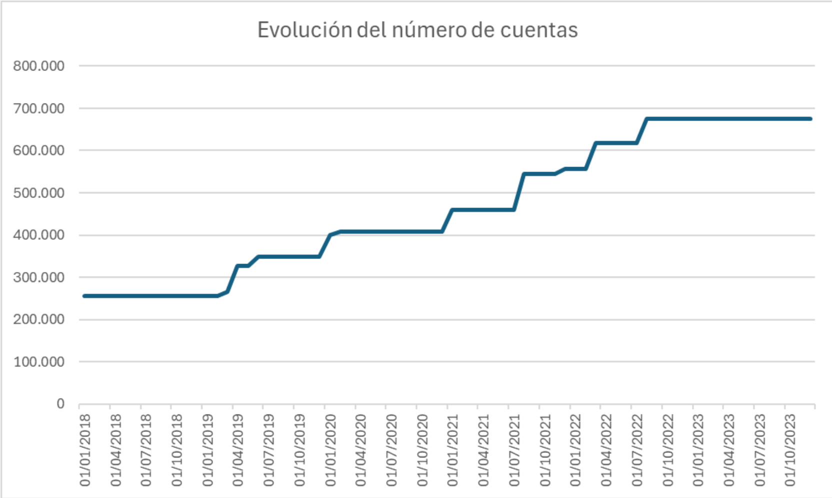
This section outlines the primary features of the data sets, along with descriptions and analyses of the key explanatory variables used in the modelling exercise.

3.4.1 Data sets for Individuals and SMEs models

The data set for Individuals comprises 676,028 distinct accounts, capturing their behaviours over the period from 2018 to 2023. This results in a comprehensive data set containing 33,472,374 entries (account-month pairings). Notably, 37.8% of these accounts are active every month analysed during this research.

Figure 3 illustrates the evolution of Individual accounts, highlighting an increase in account numbers resulting from the Investment Firm's acquisition of various portfolios. Further details about these portfolios will be discussed in the next section of this document.

Figure 3. Evolution of Individuals accounts

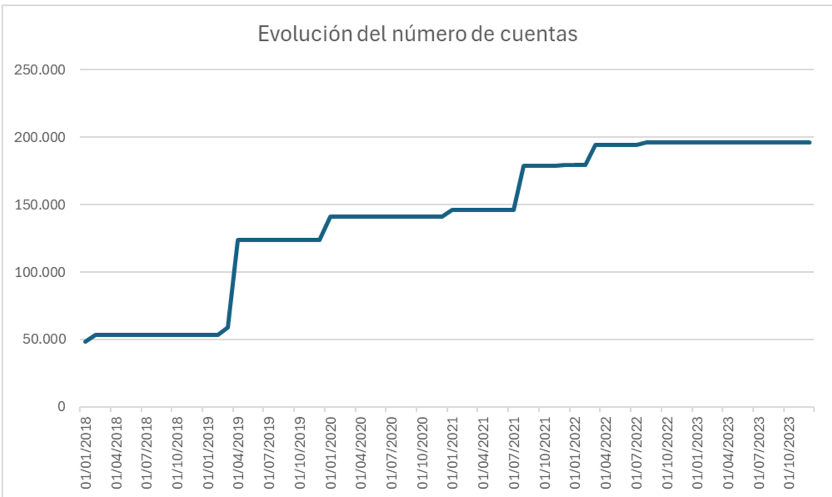


Source: Own elaboration

The dataset for SMEs comprises 195,962 distinct accounts and their behaviour from 2018 to 2023, yielding a comprehensive dataset of 10,189,744 entries (account-month pairs). Notably, 24.6% of the accounts are present in all months throughout the analysed timeframe.

Figure 4 illustrates the evolution of SME accounts, showing a steady increase following the Investment Firm's acquisition of various portfolios. The next section of this document provides further details on these portfolios.

Figure 4. Evolution of SMEs accounts



Source: Own elaboration

3.4.2 Explanatory variables

Each explanatory variable in both datasets was analysed to assess its correlation with the dependent variable and its suitability for inclusion in the constructed models. During this analysis, the information value (IV) for each variable was calculated, serving as a statistical indicator of its predictive power and helping identify the most valuable variables. IV values above 0.3 are deemed strong predictors, those between 0.1 and 0.3 are classified as medium predictors, and values below 0.1 are considered weak predictors.

Table 9 presents the IVs of the variables from the Individuals dataset, while Table 10 provides the IVs from the SMEs dataset.

In this section, we detail the top 10 variables for each dataset based on IV results. This includes a brief description of the variables, the distribution of the datasets by each variable, and a bivariate analysis of payment within the next 30 days. Appendix 3 contains details for the remaining variables.

Notably, there is a coincidence of 8 out of the top 10 variables between the Individuals and SMEs datasets, resulting in a total of 12 variables discussed in this section: Payment in the last 90 days flag, Payment ever flag, Payment in the last 30 days flag, Monthly debt amount grouped, Initial total debt amount, Product type, Initial principal debt amount, SMS sent in the last 90 days flag, Product type grouped, Promise to pay ever flag, Guarantor flag, and Judicial flag.

Additionally, this section includes a description of the macroeconomic factors and their significance for the models developed.

Table 9. Information value (IV) of variables in the Individuals data set

Variable	Description	IV Individuals
IN42_Pago_L90D	Payment in the last 90 days flag	3.73
IN43_Pago_EVER	Payment ever flag	2.86
IN41_Pago_L30D	Payment in the last 30 days flag	2.81
IN28_Saldo_InicioCorte_G	Monthly debt amount grouped	1.54
IN05_Saldo_INI_G	Initial total debt amount	1.27
IN04_Tipo_Producto	Product type	1.12
IN07_Saldo_INI_NOM_G	Initial principal debt amount	1.06
IN33_SMS_L90D	SMS sent in the last 90 days flag	0.94
IN12_Tipo_Producto_G	Product type grouped	0.80
IN40_PTP_ EVER	Promise to pay ever flag	0.73
IN39_RPC_ EVER	Right-party contacts ever flag	0.70
IN29_SMS_L30D	SMS sent in the last 30 days flag	0.68
IN37_SMS_EVER	SMS sent ever flag	0.57
IN35_RPC_L90D	Right-party contacts in the last 90 days flag	0.55
IN01_Cartera_G	Portfolio	0.47
IN27_Judicial	Judicial flag	0.45
IN03_Originador	Creditor	0.35
IN31_RPC_L30D	Right-party contacts in the last 30 days flag	0.31
IN36_PTP_L90D	Promise to pay in the last 90 days flag	0.31
IN06_Saldo_INI_INT_G	Initial interest debt amount	0.30
IN34_N_LLAMADAS_L90D	Calls in the last 90 days flag	0.27
IN09_Mes_Impago	Month & year of the original debt	0.25
IN10_Anio_Impago	Year of the original debt	0.19
IN30_N_LLAMADAS_L30D	Calls in the last 30 days flag	0.17
IN11_Originador_G	Creditor grouped	0.16
IN38_N_LLAMADAS_ EVER	Calls ever flag	0.16
IN24_NumPersonas	Number of participants	0.15
IN32_PTP_L30D	Promise to pay in the last 30 days flag	0.14
IN18_NE_NAC	National nationality flag	0.12
IN25_TotalTelefonos	Number of phone numbers	0.11
IN16_NUM_TEL_MOVIL_Sum	Number of mobile phone numbers	0.10
IN17_NE_EXT	Foreign nationality flag	0.08
IN08_Saldo_INI_OTH_G	Initial other debt amount	0.08
IN22_TP_AVA	Guarantor flag	0.08
IN13_Nacimiento_Min	Oldest date of birth or date of company creation of participants	0.07
IN15_NUM_TEL_FIJO_Sum	Number of fixed line phone numbers	0.06
IN20_Provincia	Province	0.05
IN26_CCAA	Region	0.04
IN14_Nacimiento_Max	Youngest date of birth or date of company creation of participants	0.04
IN21_TP_TIT	Holder or active company flag	0.03
IN19_NE_OTH	Other nationality flag	0.01
IN23_TP_OTH	Other participant or relationship flag	0.00

Source: Own elaboration

Table 10. Information value (IV) of variables in the SMEs data set

Variable	Description	IV SMEs
IN43_Pago_EVER	Payment ever flag	3.19
IN42_Pago_L90D	Payment in the last 90 days flag	3.04
IN41_Pago_L30D	Payment in the last 30 days flag	2.17
IN04_Tipo_Producto	Product type	1.63
IN28_Saldo_InicioCorte_G	Monthly debt amount grouped	1.39
IN05_Saldo_INI_G	Initial total debt amount	1.19
IN07_Saldo_INI_NOM_G	Initial principal debt amount	1.15
IN12_Tipo_Producto_G	Product type grouped	1.05
IN22_TP_AVA	Guarantor flag	1.04
IN27_Judicial	Judicial flag	1.02
IN13_Nacimiento_Min	Oldest date of birth or date of company creation of participants	0.81
IN21_TP_TIT	Holder or active company flag	0.56
IN14_Nacimiento_Max	Youngest date of birth or date of company creation of participants	0.52
IN01_Cartera_G	Portfolio	0.46
IN03_Originador	Creditor	0.31
IN16_NUM_TEL_MOVIL_Sum	Number of mobile phone numbers	0.30
IN25_TotalTelefonos	Number of phone numbers	0.30
IN06_Saldo_INI_INT_G	Initial interest debt amount	0.27
IN24_NumPersonas	Number of participants	0.27
IN39_RPC_ EVER	Right-party contacts ever flag	0.26
IN09_Mes_Impago	Month & year of the original debt	0.25
IN37_SMS_EVER	SMS sent ever flag	0.20
IN38_N_LLAMADAS_ EVER	Calls ever flag	0.20
IN15_NUM_TEL_FIJO_Sum	Number of fixed line phone numbers	0.19
IN11_Originador_G	Creditor grouped	0.15
IN35_RPC_L90D	Right-party contacts in the last 90 days flag	0.14
IN10_Anio_Impago	Year of the original debt	0.12
IN34_N_LLAMADAS_L90D	Calls in the last 90 days flag	0.12
IN20_Provincia	Province	0.10
IN26_CCAA	Region	0.08
IN31_RPC_L30D	Right-party contacts in the last 30 days flag	0.08
IN33_SMS_L90D	SMS sent in the last 90 days flag	0.08
IN30_N_LLAMADAS_L30D	Calls in the last 30 days flag	0.07
IN40_PTP_ EVER	Promise to pay ever flag	0.07
IN08_Saldo_INI_OTH_G	Initial other debt amount	0.05
IN29_SMS_L30D	SMS sent in the last 30 days flag	0.05
IN36_PTP_L90D	Promise to pay in the last 90 days flag	0.02
IN19_NE_OTH	Other nationality flag	0.02
IN32_PTP_L30D	Promise to pay in the last 30 days flag	0.01
IN17_NE_EXT	Foreign nationality flag	0.00
IN18_NE_NAC	National nationality flag	0.00
IN23_TP_OTH	Other participant or relationship flag	0.00

Source: Own elaboration

3.4.2.1 Variable "IN42_Pago_L90D" – Payment in the Last 90 Days Flag

Description

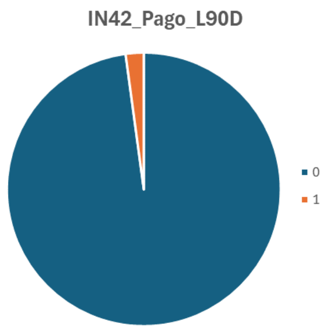
This variable indicates whether the account had a Partial Payment in the last 90 days for each month of the analysis period. For the construction of this flag, the last available data point for any given month was used. This variable was included in the data set because it typically shows a correlation with the probability of collection. If an account has a payment history, the probability of collecting is much higher than if it does not. As stated in Section 4.3 of this document, our modelling focused on the non-payer population to make this modelling exercise more useful in practice. So, this variable was not considered in the construction of our models.

Individuals' details

Analysing the last available data point for each bionomy month-account, we observe that only 2.1% of the accounts had a payment in the last ninety days.

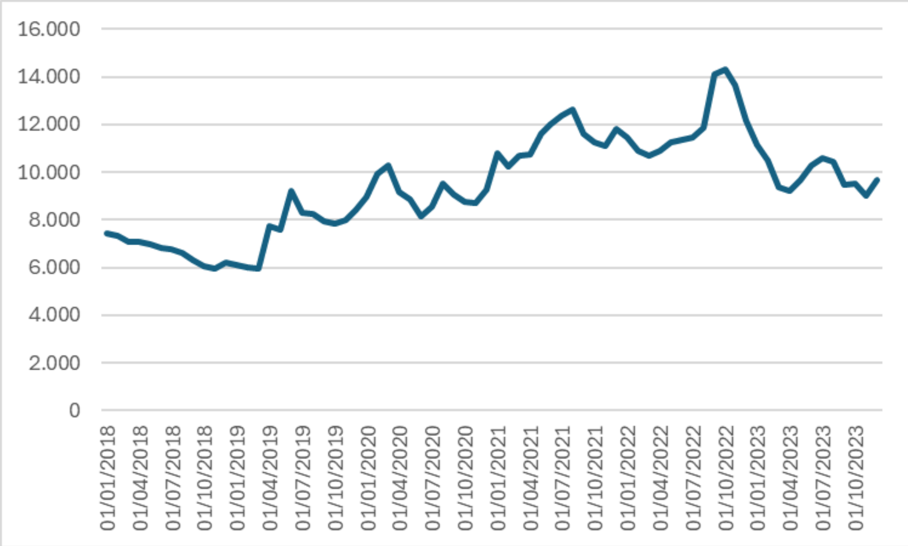
Table 11 and Figure 5. Distribution of Individuals accounts by Payment L90D flag

IN42_Pago_L90D	Record_Count
0	661.588
1	14.440



Source: Own elaboration

Figure 6. Monthly evolution of Individuals accounts by Payment L90D flag



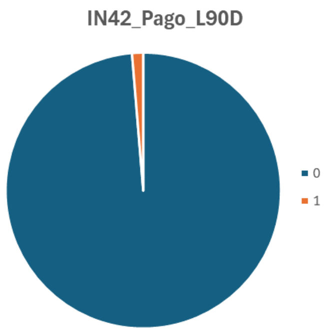
Source: Own elaboration

SMEs details

Analysing the last available data point for each bionomy month-account, we observe that 1.3% of the accounts had a payment in the last ninety days.

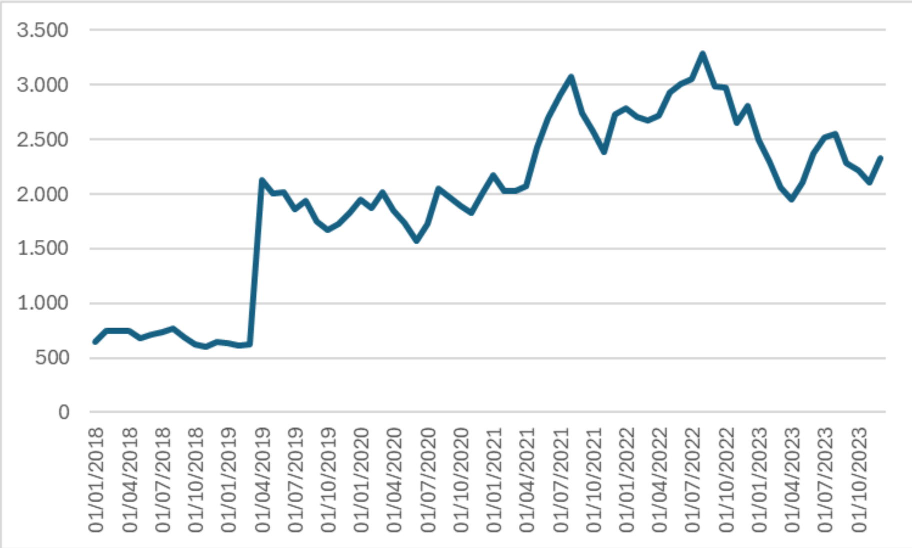
Table 12 and Figure 7. Distribution of SMEs accounts by Payment L90D flag

IN42_Pago_L90D	Record_Count
0	193.389
1	2.573



Source: Own elaboration

Figure 8. Monthly evolution of SMEs accounts by Payment L90D flag



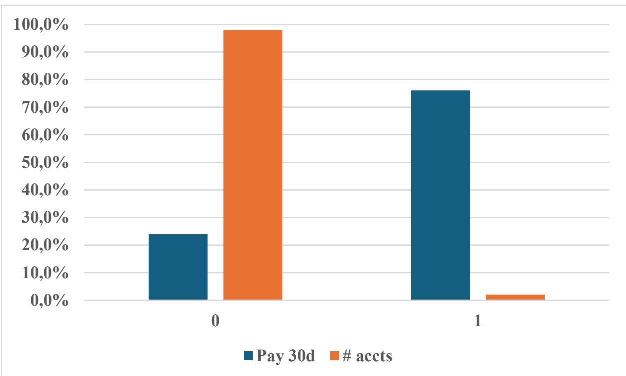
Source: Own elaboration

Discrimination Bivariate Analysis for Individuals and SMEs

In this variable, we observe a very high correlation between the payment in the last 90 days flag and the payment in the following 30 days (the distributions of accounts and payments differ between positive and negative flag outcomes). In both Individuals and SMEs, the probability of payment is much higher when the flag result is positive. The information value (IV) of this variable is 3.73 for Individuals and 3.04 for SMEs.

Table 13 and Figure 9. Bivariate analysis payment last 90d flag – payment 30 days for Individuals

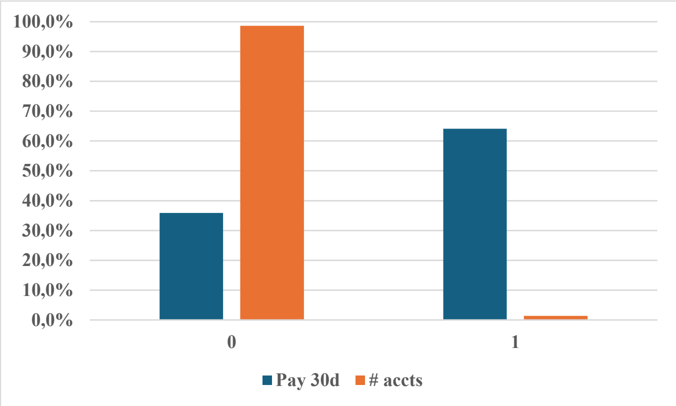
Payment last 90d flag	Pay 30d	# accts
0	23,9%	98,0%
1	76,1%	2,0%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 14 and Figure 10. Bivariate analysis payment last 90d flag – payment 30 days for SMEs

Payment last 90d flag	Pay 30d	# accts
0	35,9%	98,6%
1	64,1%	1,4%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.4.2.2 Variable "IN43_Pago_EVER" – Payment Ever Flag

Description

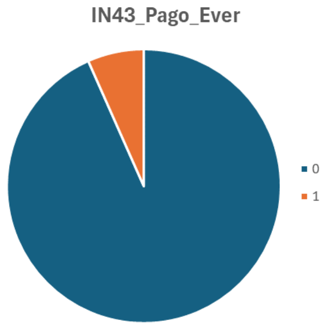
This variable indicates whether the account had a previous Partial Payment in each month of the analysis period. For the construction of this flag, the last available data point for any given month was used. This variable was included in the data set because it typically shows a correlation with the probability of collection. If an account has a payment history, the probability of collecting is much higher than if it does not. As stated in Section 4.3 of this document, our modelling focused on the non-payer population to make the exercise more useful in practice. So, this variable was not considered in the construction of our models.

Individuals' details

Analysing the last available data point for each bionomy month-account, we observe that 6.6% of the accounts had a payment ever (in the analysis period).

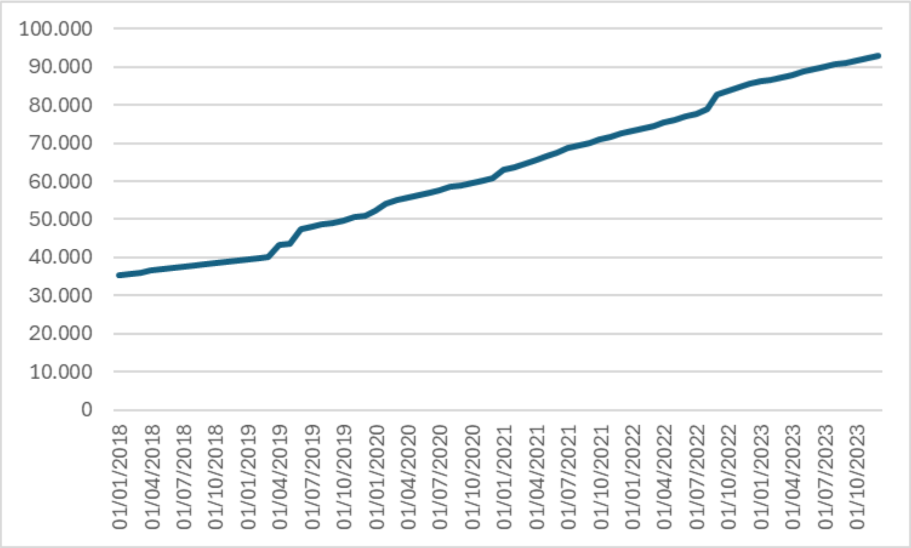
Table 15 and Figure 11. Distribution of Individuals accounts by Payment Ever flag

IN43_Pago_Ever	Record_Count
0	631.109
1	44.919



Source: Own elaboration

Figure 12. Monthly evolution of Individuals accounts by Payment Ever flag



Source: Own elaboration

SMEs details

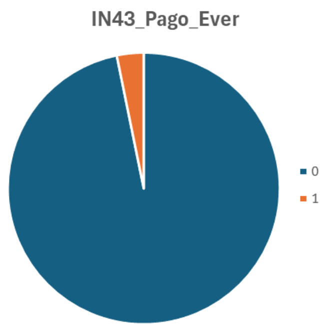
Analysing the last available data point for each bionomy month-account, we observe that 3.2% of the accounts had a payment ever (in the analysis period).

Table 16. Distribution of SMEs accounts by Payment Ever flag

IN43_Pago_Ever	Record_Count
0	189.704
1	6.258

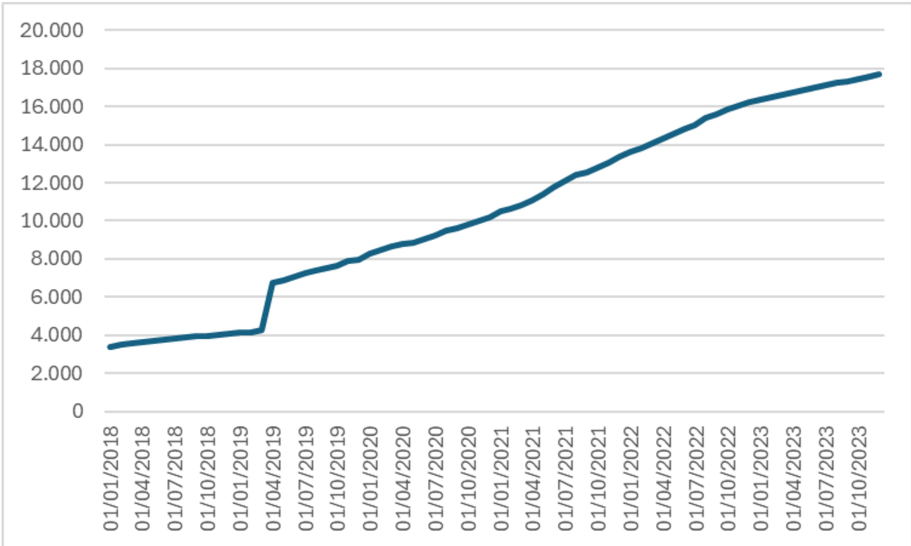
Source: Own elaboration

Figure 13. Distribution of SMEs accounts by Payment Ever flag



Source: Own elaboration

Figure 14. Monthly evolution of SMEs accounts by Payment Ever flag



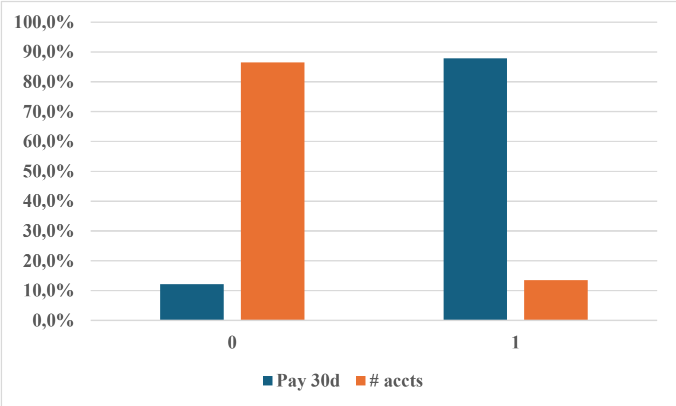
Source: Own elaboration

Discrimination Bivariate Analysis for Individuals and SMEs

In this variable, we observe a very high correlation between the payment ever flag and payments in the following 30 days (the distributions of accounts and payments differ between positive and negative flag outcomes). In both Individuals and SMEs, the probability of payment is much higher when the flag result is positive. The information value (IV) of this variable is 2.86 for Individuals and 3.19 for SMEs.

Table 17 and Figure 15. Bivariate analysis payment ever flag – payment 30 days for Individuals

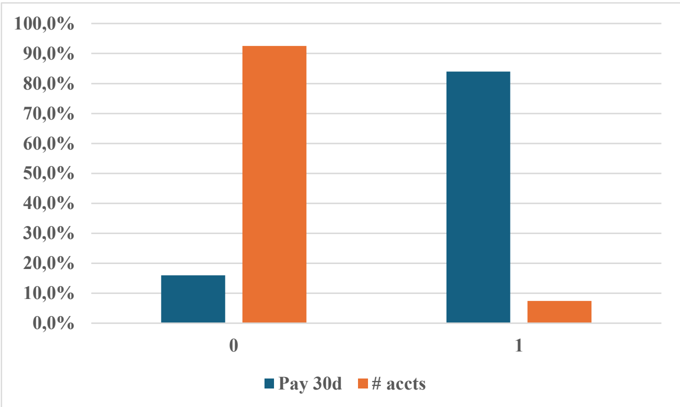
Payment ever flag	Pay 30d	# accts
0	12,1%	86,5%
1	87,9%	13,5%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 18 and Figure 16. Bivariate analysis payment ever flag – payment 30 days for SMEs

Payment ever flag	Pay 30d	# accts
0	16,0%	92,5%
1	84,0%	7,5%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.4.2.3 Variable "IN41_Pago_L30D" – Payment in the Last 30 Days Flag

Description

This variable indicates whether the account had a Partial Payment in the last 30 days for each month of the analysis period. For the construction of this flag, the last available data point for any given month was used. This variable was included in the data set because it typically shows a correlation with the probability of collection. If an account has a payment history, the probability of collecting is much higher than if it does not. As stated in Section 4.3 of this document, our modelling focused on the non-payer population to make this modelling exercise more useful in practice. So, this variable was not considered in the construction of our models.

Individuals' details

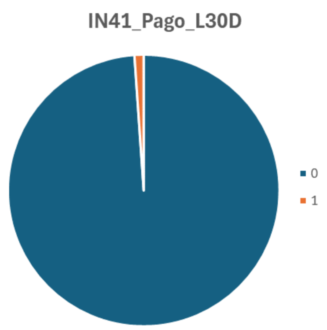
Analysing the last available data point for each bionomy month-account, we observe that only 1.1% of the accounts had a payment in the last thirty days.

Table 19. Distribution of Individuals accounts by Payment L30D flag

IN41_Pago_L30D	Record_Count
0	668.629
1	7.399

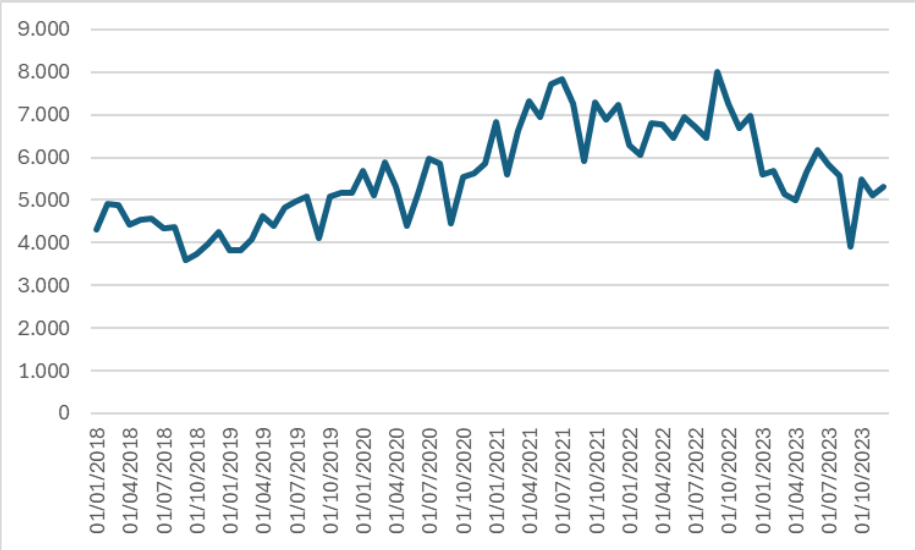
Source: Own elaboration

Figure 17. Distribution of Individuals accounts by Payment L30D flag



Source: Own elaboration

Figure 18. Monthly evolution of Individuals accounts by Payment L30D flag



Source: Own elaboration

SMEs details

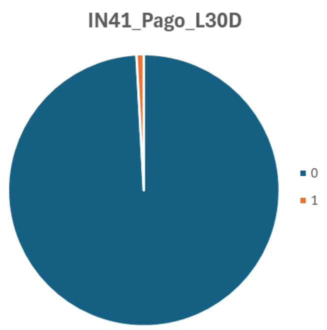
Analysing the last available data point for each bionomy month-account, we observe that only 0.9% of the accounts had a payment in the last thirty days.

Table 20. Distribution of SMEs accounts by Payment L30D flag

IN41_Pago_L30D	Record_Count
0	194.270
1	1.692

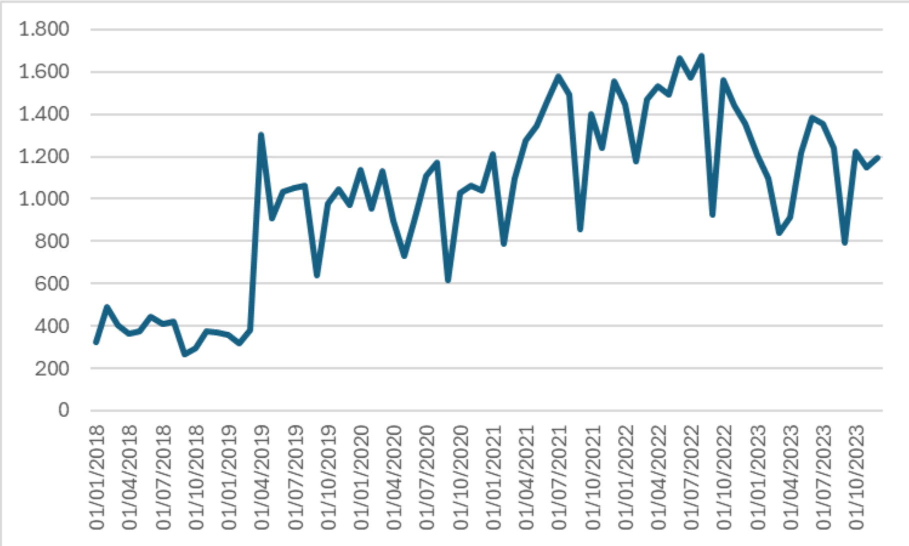
Source: Own elaboration

Figure 19. Distribution of SMEs accounts by Payment L30D flag



Source: Own elaboration

Figure 20. Monthly evolution of SMEs accounts by Payment L30D flag



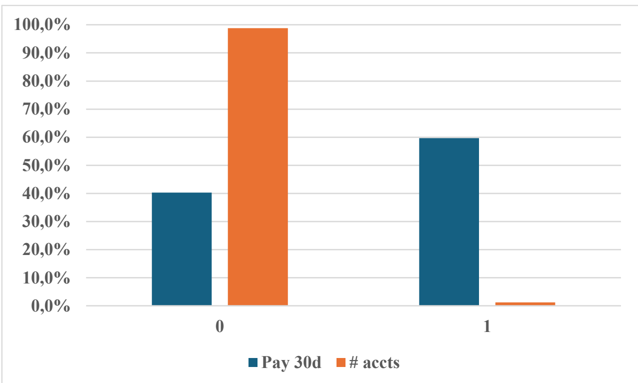
Source: Own elaboration

Discrimination Bivariate Analysis for Individuals and SMEs

In this variable, we observe a very high correlation between the payment last 30 days flag and the payment in the following 30 days (the distributions of accounts and payments differ between positive and negative flag outcomes). In both Individuals and SMEs, the probability of payment is much higher when the flag result is positive. The information value (IV) of this variable is 2.81 for Individuals and 2.17 for SMEs.

Table 21 and Figure 21. Bivariate analysis payment last 30d flag – payment 30 days for Individuals

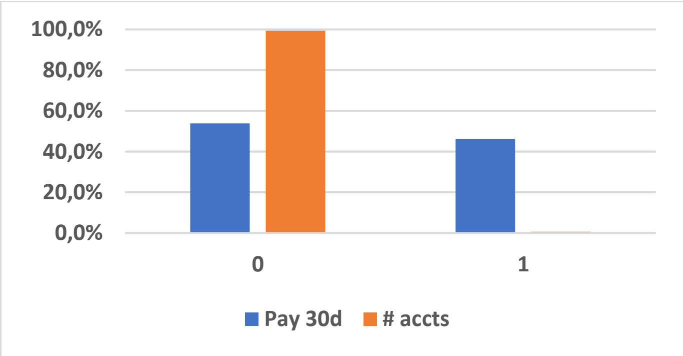
Payment last 30d flag	Pay 30d	# accts
0	40,3%	98,8%
1	59,7%	1,2%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 22 and Figure 22. Bivariate analysis payment last 30d flag – payment 30 days for SMEs

Payment last 30d flag	Pay 30d	# accts
0	53,8%	99,3%
1	46,2%	0,7%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.4.2.4 Variable "IN28_Saldo_InicioCorte_G" – Monthly Debt Amount Grouped

Description

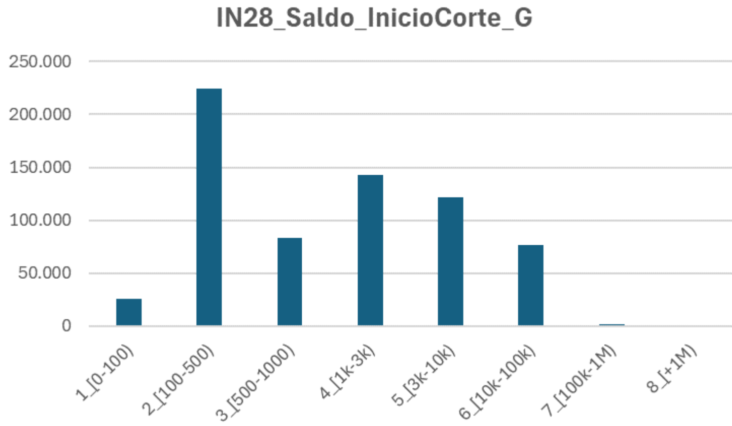
This variable reflects the monthly debt amount for each account during the analysis period, grouped into several ranges. This variable was included in the data set because it typically shows a correlation with the probability of collection. Lower amounts indicate a higher probability of collection, and higher amounts indicate a lower probability. As is common in practice, this variable was relevant to several of the models we constructed: Individuals COVID, Individuals Post-COVID, SMEs Global, SMEs Pre-COVID, and SMEs Post-COVID.

Individuals' details

Analysing the last available data point for each bionomy month-account, we observe that 37% of the total accounts have a debt amount below 500€, and 12% have a debt amount above 10k€.

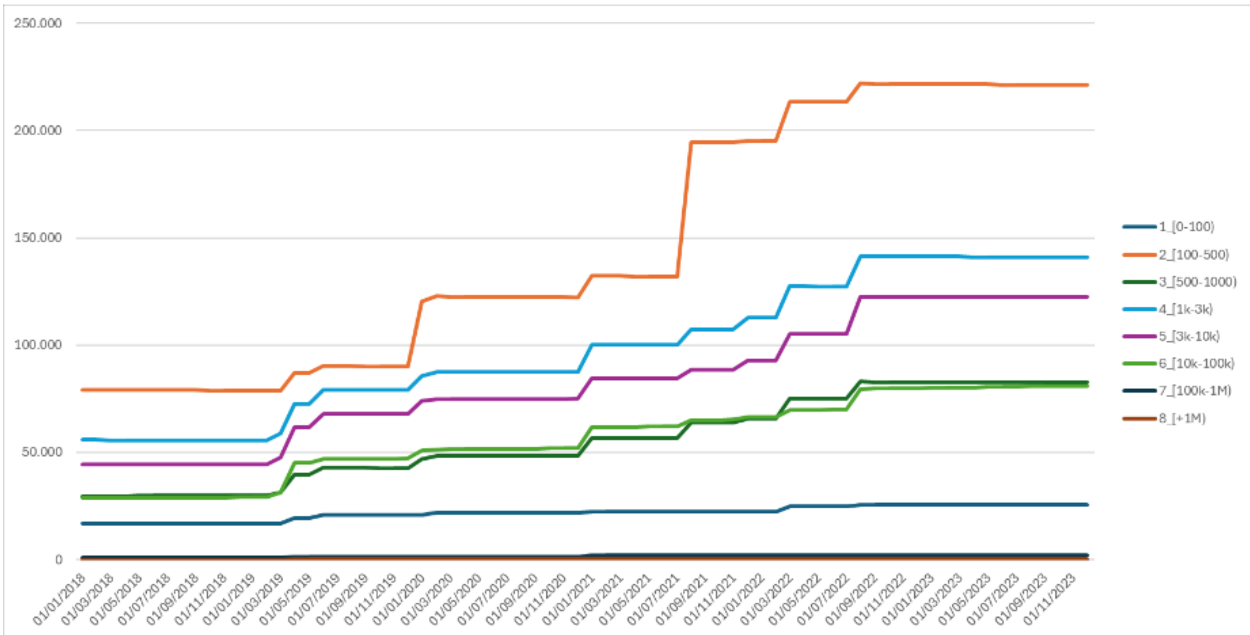
Table 23 and Figure 23. Distribution of Individuals accounts by monthly debt amount grouped

IN28_Saldo_InicioCorte_G	Record_Count
1_[0-100)	26.113
2_[100-500)	224.170
3_[500-1000)	82.952
4_[1k-3k)	142.497
5_[3k-10k)	121.909
6_[10k-100k)	76.317
7_[100k-1M)	2.031
8_[+1M)	39



Source: Own elaboration

Figure 24. Monthly evolution of Individuals accounts by debt amount grouped



Source: Own elaboration

SMEs details

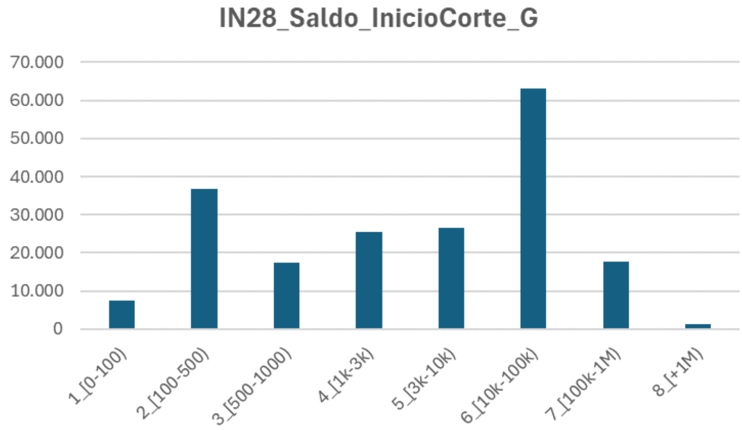
Analysing the last available data point for each bionomy month-account, we observe that 23% of the total accounts have a debt amount below 500€, and 42% have a debt amount above 10k€.

Table 24. Distribution of SMEs accounts by monthly debt amount grouped

IN28_Saldo_InicioCorte_G	Record_Count
1_[0-100)	7.484
2_[100-500)	36.863
3_[500-1000)	17.459
4_[1k-3k)	25.597
5_[3k-10k)	26.579
6_[10k-100k)	63.002
7_[100k-1M)	17.695
8_[+1M)	1.283

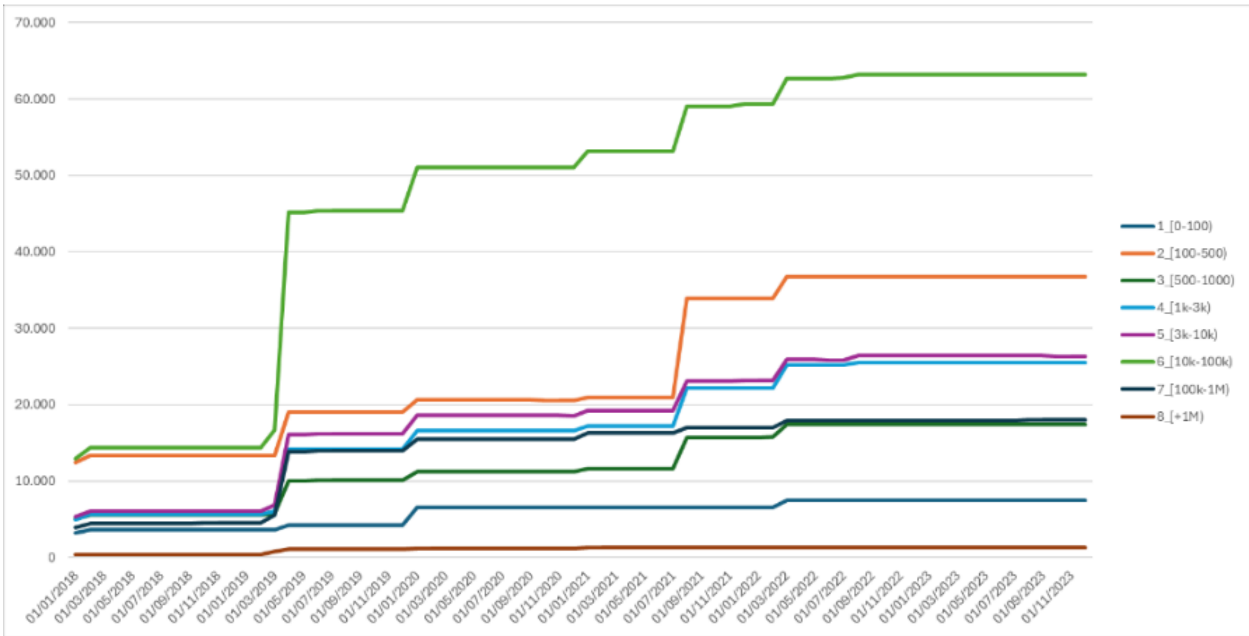
Source: Own elaboration

Figure 25. Distribution of SMEs accounts by monthly debt amount grouped



Source: Own elaboration

Figure 26. Monthly evolution of SMEs accounts by debt amount grouped



Source: Own elaboration

Discrimination Bivariate Analysis for Individuals and SMEs

For this variable, we observe a correlation between monthly debt amounts grouped and payments in the following 30 days (the distributions of accounts and payments differ across monthly debt amount ranges). In Individuals, the probability for a payment is higher in the ranges above 3k. For SMEs, this probability is higher for ranges above 10k. The information value (IV) of this variable is 1.54 for Individuals and 1.39 for SMEs.

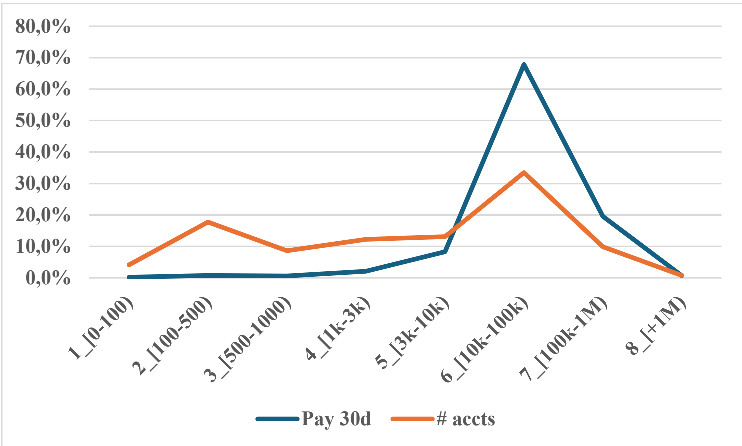
Table 25 and Figure 27. Bivariate analysis monthly debt amount grouped – payment 30 days for Individuals

Monthly debt amount grouped	Pay 30d	# accts
1_[0-100)	0,3%	4,7%
2_[100-500)	2,9%	31,6%
3_[500-1000)	2,9%	12,1%
4_[1k-3k)	13,1%	21,2%
5_[3k-10k)	33,0%	17,9%
6_[10k-100k)	46,9%	12,2%
7_[100k-1M)	0,9%	0,3%
8_[+1M)	0,0%	0,0%
Total general	**100,0%**	**100,0%**

Source: Own elaboration

Table 26 and Figure 28. Bivariate analysis monthly debt amount grouped – payment 30 days for SMEs

Monthly debt amount grouped	Pay 30d	# accts
1_[0-100)	0,2%	4,2%
2_[100-500)	0,7%	17,8%
3_[500-1000)	0,6%	8,7%
4_[1k-3k)	2,1%	12,2%
5_[3k-10k)	8,3%	13,1%
6_[10k-100k)	67,8%	33,5%
7_[100k-1M)	19,5%	9,9%
8_[+1M)	0,7%	0,7%
Total general	100,0%	100,0%



Source: Own elaboration

3.4.2.5 Variable "IN05_Saldo_INI_G" – Initial Total Debt Amount

Description

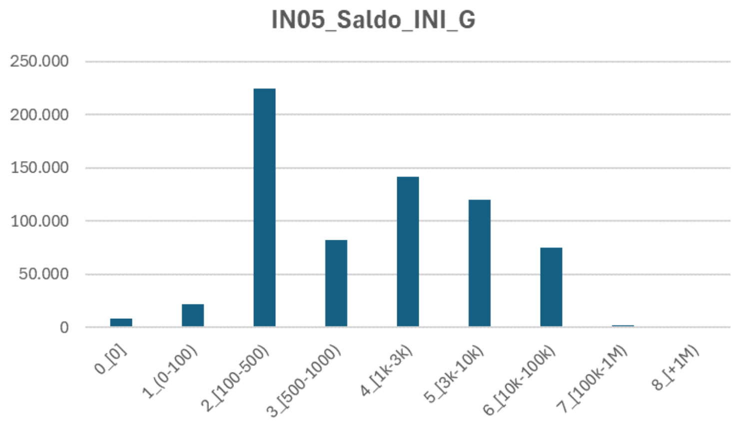
Initial total outstanding debt amount of the account, grouped in ranges or buckets. The bucket distribution defined for this variable will be applied to the remaining variables linked to the debt amount. This variable was included in the data set because it typically shows a correlation with the probability of collection. Lower amounts indicate a higher probability of collection, and higher amounts indicate a lower probability. This variable was relevant for our modelling only in the Individuals Pre-COVID and SMEs Pre-COVID models.

Individuals' details

There is a uniform distribution of accounts across the defined buckets in the dataset. The bucket with the most accounts is the 100-500€ range of total initial debt, representing 33% of the total accounts. There is a low presence of accounts with total initial debt exceeding 100k€ (0.3% of total accounts), and 1.3% of accounts lack information or have null values for total initial debt.

Table 27 and Figure 29. Distribution of Individuals accounts by initial total debt amount

IN05_Saldo_INI_G	Record_Count
0_[0]	8.498
1_(0-100)	22.193
2_[100-500)	224.595
3_[500-1000)	81.999
4_[1k-3k)	141.757
5_[3k-10k)	120.153
6_[10k-100k)	74.790
7_[100k-1M)	2.004
8_[+1M)	39



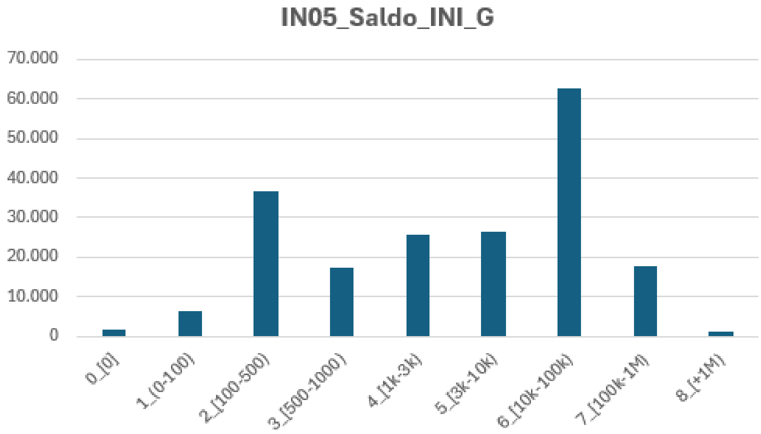
Source: Own elaboration

SMEs details

There is a uniform distribution of accounts across the defined buckets in the dataset. The bucket with the most accounts is the 10k-100k€ range of total initial debt, representing 32% of the total accounts. There is a low presence of accounts with total initial debt exceeding 1M€ (0.7% of total accounts), and 0.9% of total accounts lack information or have null values for total initial debt.

Table 28 and Figure 30. Distribution of SMEs accounts by initial total debt amount

IN05_Saldo_INI_G	Record_Count
0_[0]	1.703
1_(0-100)	6.177
2_[100-500)	36.828
3_[500-1000)	17.427
4_[1k-3k)	25.554
5_[3k-10k)	26.524
6_[10k-100k)	62.847
7_[100k-1M)	17.628
8_[+1M)	1.274



Source: Own elaboration

Discrimination Bivariate Analysis for Individuals and SMEs

In this case, we observe a correlation between the initial total debt amount and payments in the following 30 days (the distribution of accounts and payments differs across different ranges of the initial total debt amount). For Individuals, the probability of a payment is higher in the ranges of 3-10k, 10-100k, and 100k-1M. Moreover, for SMEs, the probability is higher in the 10-100k and 100k-1M ranges. The information value (IV) of this variable is 1.27 for Individuals and 1.19 for SMEs.

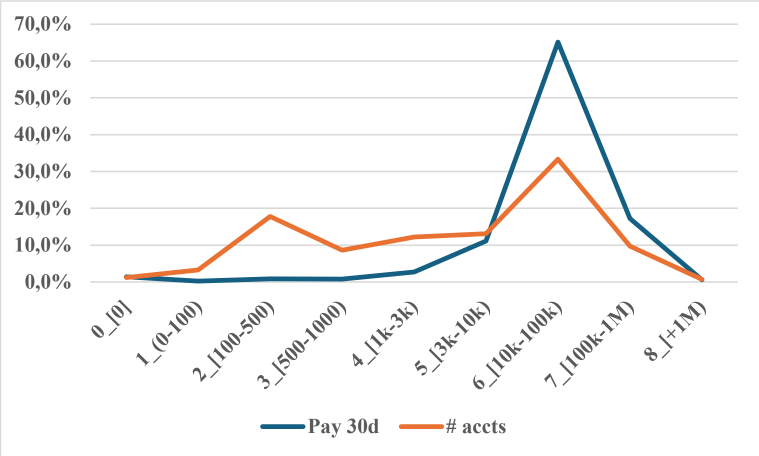
Table 29 and Figure 31. Bivariate analysis initial total debt amount – payment 30 days for Individuals

Initial total debt amount	Pay 30d	# accts
0_[0]	0,1%	1,8%
1_(0-100)	0,3%	3,9%
2_[100-500)	3,9%	32,0%
3_[500-1000)	3,9%	11,9%
4_[1k-3k)	17,5%	21,2%
5_[3k-10k)	35,7%	17,5%
6_[10k-100k)	37,8%	11,5%
7_[100k-1M)	0,7%	0,3%
8_[+1M)	0,0%	0,0%
Total general	100,0%	100,0%

Source: Own elaboration

Table 30 and Figure 32. Bivariate analysis initial total debt amount – payment 30 days for SMEs

Initial total debt amount	Pay 30d	# accts
0_[0]	1,4%	1,2%
1_(0-100)	0,2%	3,3%
2_[100-500)	0,8%	17,8%
3_[500-1000)	0,8%	8,6%
4_[1k-3k)	2,7%	12,2%
5_[3k-10k)	11,1%	13,1%
6_[10k-100k)	65,1%	33,3%
7_[100k-1M)	17,2%	9,7%
8_[+1M)	0,6%	0,7%
Total general	100,0%	100,0%



Source: Own elaboration

3.4.2.6 Variable "IN04_Tipo_Producto" – Product Type

Description

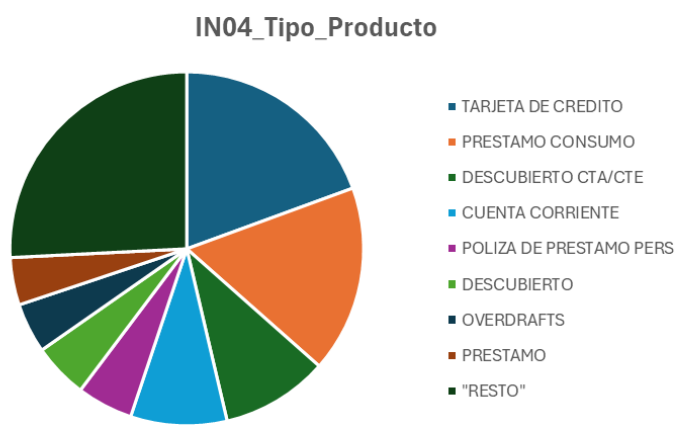
Type of product of the original debt (without grouping and before the cleaning of null cases). This variable was included in the data set because it typically shows a correlation with the probability of collection, as it is linked to the debtor profile. This variable was not relevant in our modelling, but its grouping (variable "IN12_Tipo_Producto_G" – Product type grouped) was relevant in all the constructed models.

Individuals' details

263 different product types were originally in the data set. The most significant product type is "Tarjetas de Crédito" – credit cards, and represents 19% of the total accounts.

Table 31 and Figure 33. Distribution of Individuals accounts by product type

IN04_Tipo_Producto	Record_Count
TARJETA DE CREDITO	131.335
PRESTAMO CONSUMO	115.599
DESCUBIERTO CTA/CTE	66.262
CUENTA CORRIENTE	59.622
POLIZA DE PRESTAMO PERS	34.663
DESCUBIERTO	33.992
OVERDRAFTS	30.891
PRESTAMO	29.310
TARJETA DE CREDITO / CAJA MADRID / BANKI	19.378



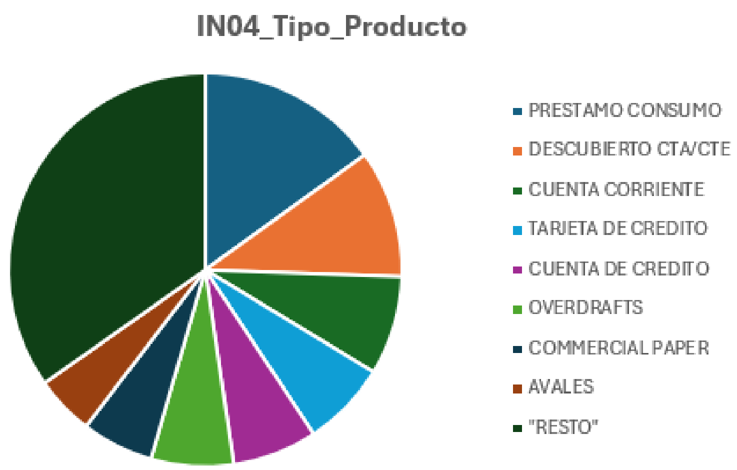
Source: Own elaboration

SMEs details

263 different product types were originally in the data set. The most significant product type is "Préstamos al consumo" – consumer loans, and represents 15% of the total accounts.

Table 32 and Figure 34. Distribution of SMEs accounts by product type

IN04_Tipo_Producto	Record_Count
PRESTAMO CONSUMO	29.534
DESCUBIERTO CTA/CTE	20.531
CUENTA CORRIENTE	15.938
TARJETA DE CREDITO	13.770
CUENTA DE CREDITO	13.662
OVERDRAFTS	13.280
COMMERCIAL PAPER	11.595
AVALES	9.543
POLIZA DE PRESTAMO PERS	7.687

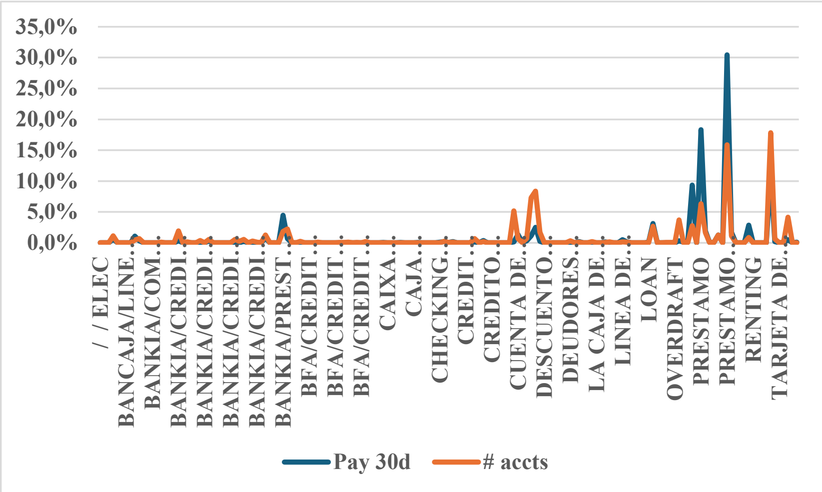


Source: Own elaboration

Discrimination Bivariate Analysis for Individuals and SMEs

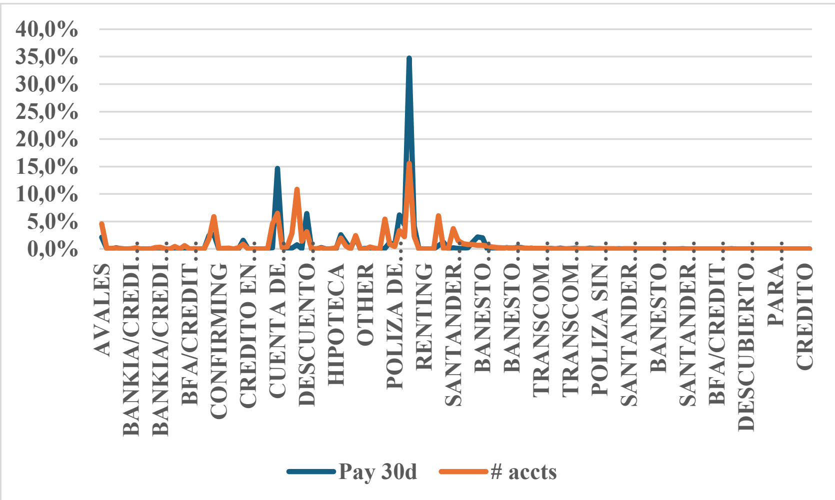
We observe a correlation between product type and payments in the following 30 days (the distributions of accounts and payments differ across product types). Certain product types have a higher probability of payment than others, both for Individuals and SMEs. In the grouped version of this variable (variable "IN12_Tipo_Producto_G" – Product type grouped), we detail which of these product types have a higher probability. The information value (IV) of this variable is 1.12 for Individuals and 1.63 for SMEs.

Figure 35. Bivariate analysis product type – payment 30 days for Individuals



Source: Own elaboration

Figure 36. Bivariate analysis product type – payment 30 days for SMEs



Source: Own elaboration

3.4.2.7 Variable "IN07_Saldo_INI_NOM_G" – Initial Principal Debt Amount

Description

The initial principal outstanding debt amount of the account, grouped in ranges or buckets. This variable was included in the data set because it typically shows a correlation with the probability of collection. Lower amounts indicate a higher probability of collection, and higher amounts indicate a lower probability. This variable was relevant in all the Individuals models and in the SMEs Global and SMEs COVID models.

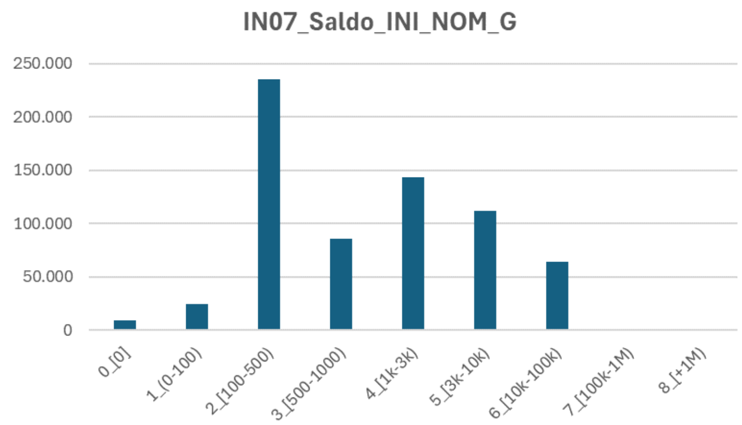
The distribution of accounts by this variable is very similar to the distribution of the variable "IN05_Saldo_INI_G" – initial total debt amount. This is mainly due to the important lack of accounts with interests or other amounts different from the principal amount.

Individuals' details

The bucket with the most accounts is the 100-500€ range of total initial debt, representing 35% of the total accounts. There is a low presence of accounts with total initial debt exceeding 100k€ (0.2% of total accounts), and 1.3% of accounts lack information or have null values for total initial debt.

Table 33 and Figure 37. Distribution of Individuals accounts by initial principal debt amount

IN07_Saldo_INI_NOM_G	Record_Count
0_[0]	8.818
1_(0-100)	24.389
2_[100-500)	235.492
3_[500-1000)	85.911
4_[1k-3k)	143.881
5_[3k-10k)	111.865
6_[10k-100k)	64.204
7_[100k-1M)	1.436
8_[+1M)	32



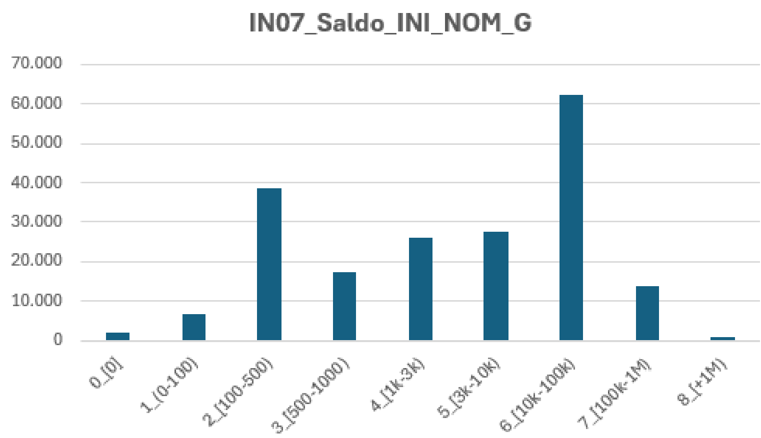
Source: Own elaboration

SMEs details

The bucket with the most accounts is the 10k-100k€ range of total initial debt, representing 32% of the total accounts. There is a low presence of accounts with total initial debt exceeding 1M€ (0.5% of total accounts), and 1.0% of total accounts lack information or have null values for total initial debt.

Table 34 and Figure 38. Distribution of SMEs accounts by initial principal debt amount

IN07_Saldo_INI_NOM_G	Record_Count
0_[0]	1.907
1_(0-100)	6.725
2_[100-500)	38.500
3_[500-1000)	17.499
4_[1k-3k)	26.153
5_[3k-10k)	27.763
6_[10k-100k)	62.487
7_[100k-1M)	13.995
8_[+1M)	933



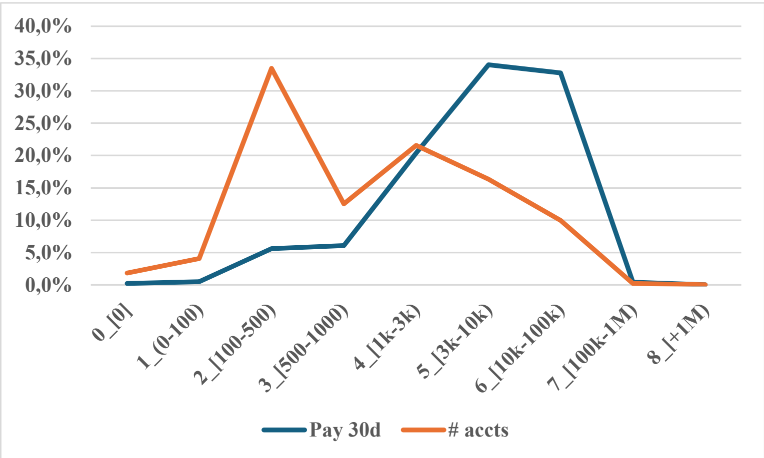
Source: Own elaboration

Discrimination Bivariate Analysis for Individuals and SMEs

In this case, we observe a correlation between the initial principal debt amount and payments in the following 30 days (the distribution of accounts and payments differs across different ranges of the initial principal debt amount). In Individuals, the probability of a payment is higher across all ranges above 3k. Furthermore, in SMEs, the probability is higher in the 3-10k, 10-100k, and 100k-1M ranges. The information value (IV) of this variable is 1.06 for Individuals and 1.15 for SMEs.

Table 35 and Figure 39. Bivariate analysis initial principal debt amount – payment 30 days for Individuals

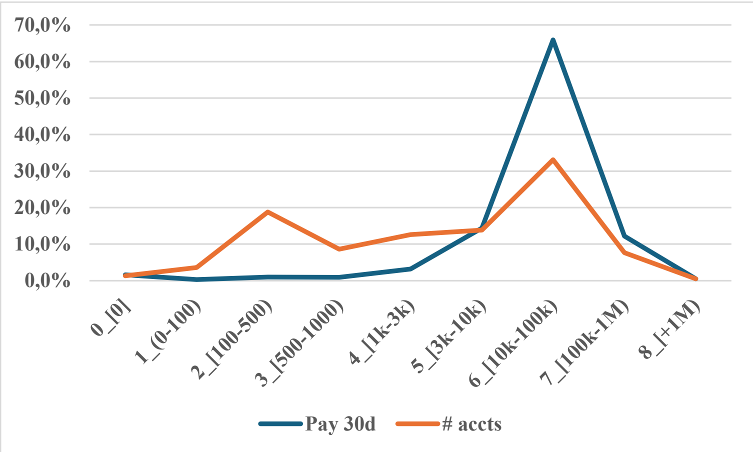
Initial principal debt amount	Pay 30d	# accts
0_[0]	0,2%	1,8%
1_(0-100)	0,5%	4,1%
2_[100-500)	5,6%	33,5%
3_[500-1000)	6,1%	12,5%
4_[1k-3k)	20,4%	21,6%
5_[3k-10k)	34,0%	16,4%
6_[10k-100k)	32,8%	10,0%
7_[100k-1M)	0,4%	0,2%
8_[+1M)	0,0%	0,0%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 36 and Figure 40. Bivariate analysis initial principal debt amount – payment 30 days for SMEs

Initial principal debt amount	Pay 30d	# accts
0_[0]	1,6%	1,3%
1_(0-100)	0,3%	3,6%
2_[100-500)	1,0%	18,8%
3_[500-1000)	0,9%	8,6%
4_[1k-3k)	3,2%	12,6%
5_[3k-10k)	14,4%	13,8%
6_[10k-100k)	66,0%	33,1%
7_[100k-1M)	12,2%	7,6%
8_[+1M)	0,5%	0,5%
Total general	100,0%	100,0%



Source: Own elaboration

3.4.2.8 Variable "IN33_SMS_L90D" – SMS Sent in the Last 90 Days Flag

Description

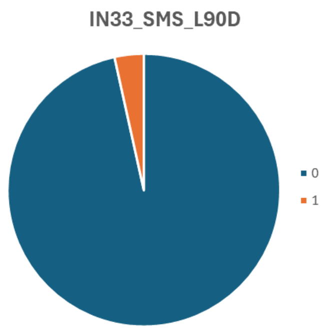
This variable indicates whether an SMS was sent to the account in the last 90 days for each month of the analysis period. For the construction of this flag, the last available data point for any given month was used. This variable was included in the data set because it typically shows a correlation with the probability of collection. The collection and recovery actions (SMS, calls…) are positively correlated with the probability of collection: the more collection actions, the higher the probability of collection. As explained in Section 4.4.3 of this document, this variable was excluded from our modelling due to excessive correlation with the predicted variable.

Individuals' details

Analysing the last available data point for each bionomy month-account, we observe that just 3.5% of the accounts have an SMS sent in the last 90 days.

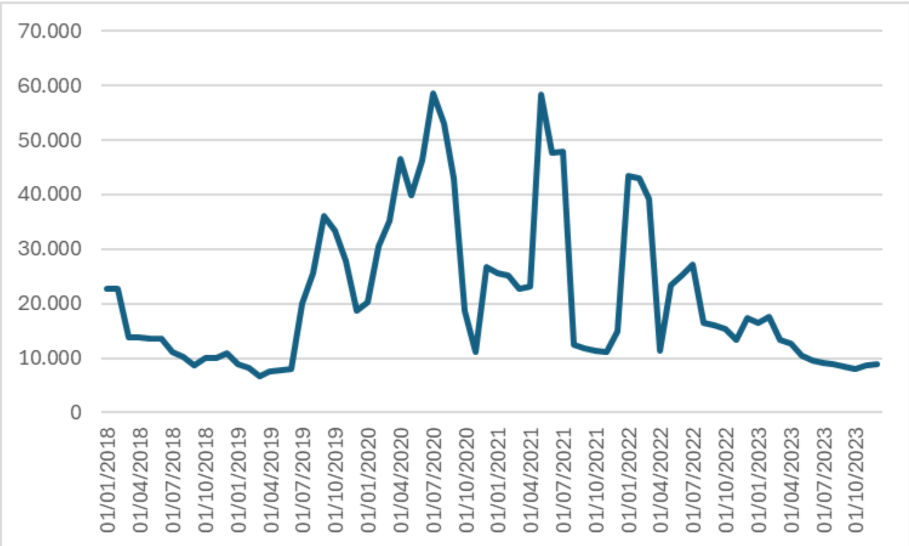
Table 37 and Figure 41. Distribution of Individuals accounts by SMS L90D flag

IN33_SMS_L90D	Record_Count
0	652.602
1	23.426



Source: Own elaboration

Figure 42. Monthly evolution of Individuals accounts by SMS L90D flag



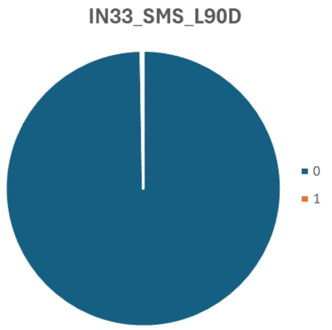
Source: Own elaboration

SMEs details

Analysing the last available data point for each bionomy month-account, we observe that just 0.3% of the accounts have an SMS sent in the last 90 days.

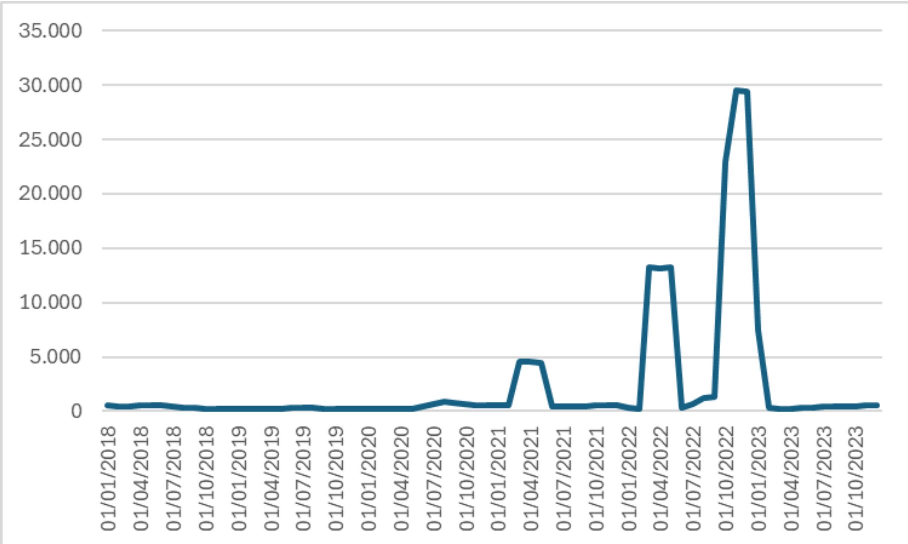
Table 38 and Figure 43. Distribution of SMEs accounts by SMS L90D flag

IN33_SMS_L90D	Record_Count
0	195.417
1	545



Source: Own elaboration

Figure 44. Monthly evolution of SMEs accounts by SMS L90D flag



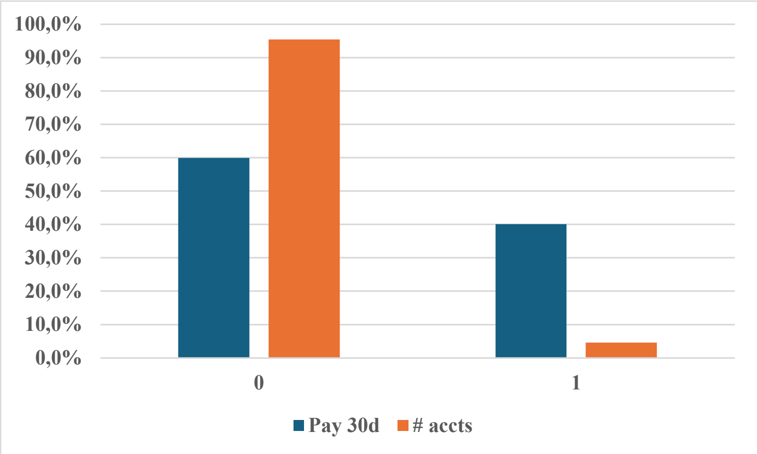
Source: Own elaboration

Discrimination Bivariate Analysis for Individuals and SMEs

In this variable, we observe a strong correlation between the SMS sent in the last 90 days flag and payments in the following 30 days (the distributions of accounts and payments differ between positive and negative flag outcomes). In both Individuals and SMEs, the probability of payment is much higher when the flag result is positive. The information value (IV) of this variable is 0.94 for Individuals and 0.08 for SMEs.

Table 39 and Figure 45. Bivariate analysis SMS sent last 90d flag – payment 30 days for Individuals

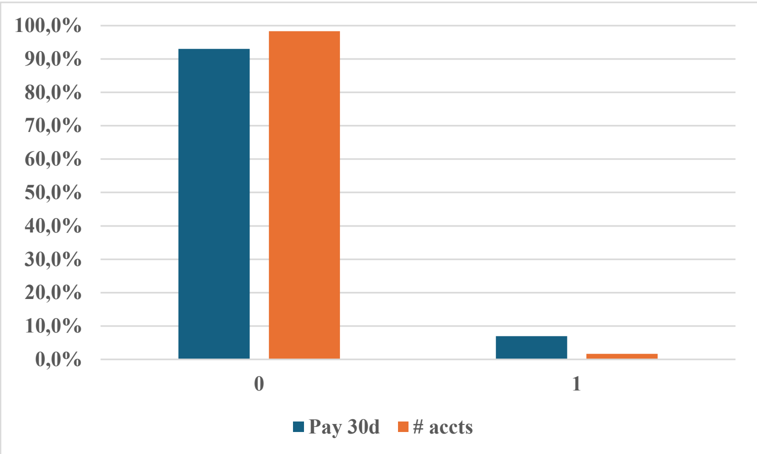
SMS sent last 90d flag	Pay 30d	# accts
0	59,9%	95,4%
1	40,1%	4,6%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 40 and Figure 46. Bivariate analysis SMS sent last 90d flag – payment 30 days for SMEs

SMS sent last 90d flag	Pay 30d	# accts
0	93,1%	98,3%
1	6,9%	1,7%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.4.2.9 Variable "IN12_Tipo_Producto_G" – Product Type Grouped

Description

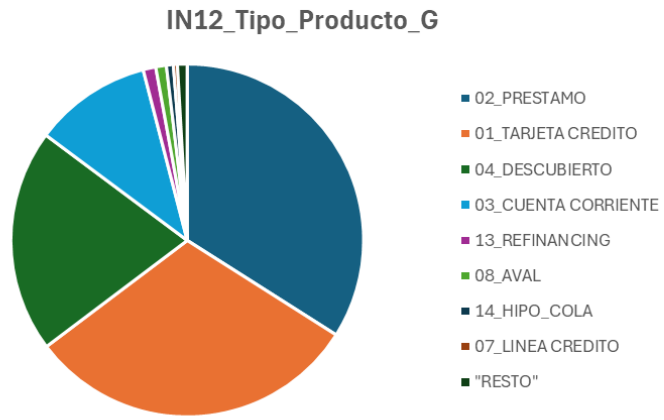
Type of product of the original debt, grouped and after cleaning of null cases. This variable was included in the data set because it typically shows a correlation with the probability of collection, as it is linked to the debtor profile. In our modelling, this variable was relevant for all the Individuals and SMEs models.

Individuals' details

15 distinct product type groups were defined in the dataset. High concentration in the loan product, with more than 34% of the total accounts. Current account, overdrafts, and credit cards represents around 62% of the total accounts.

Table 41 and Figure 47. Distribution of Individuals accounts by product type grouped

IN12_Tipo_Producto_G	Record_Count
02_PRESTAMO	229.821
01_TARJETA CREDITO	207.723
04_DESCUBIERTO	138.354
03_CUENTA CORRIENTE	72.913
13_REFINANCING	7.833
08_AVAL	6.576
14_HIPO_COLA	4.280
07_LINEA CREDITO	2.503
15_OTROS	2.423
05_COMMERCIAL PAPER	1.786
10_CREDITO	664
09_DESCUENTO	545
12_GARANTIA	490
11_RENTING/LEASING	117



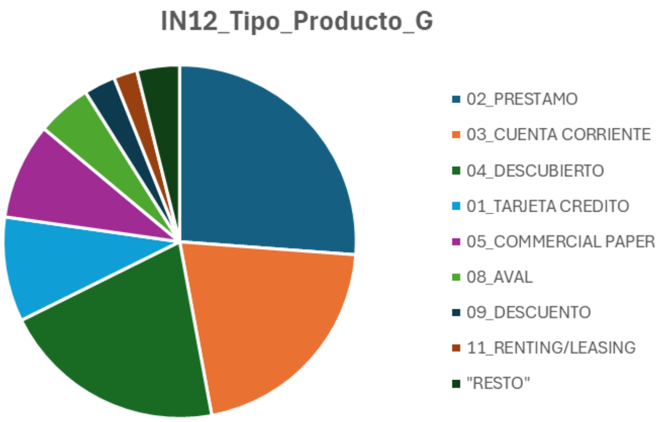
Source: Own elaboration

SMEs details

15 distinct product type groups were defined in the dataset. High concentration in the loan product, with more than 25% of the total accounts. Current account, overdrafts, and credit cards represents around 50% of the total accounts.

Table 42 and Figure 48. Distribution of SMEs accounts by product type grouped

IN12_Tipo_Producto_G	Record_Count
02_PRESTAMO	51.288
03_CUENTA CORRIENTE	40.981
04_DESCUBIERTO	40.351
01_TARJETA CREDITO	18.767
05_COMMERCIAL PAPER	17.252
08_AVAL	9.790
09_DESCUENTO	5.658
11_RENTING/LEASING	4.230
15_OTROS	2.762
07_LINEA CREDITO	2.558
10_CREDITO	1.304
14_HIPO_COLA	652
12_GARANTIA	369



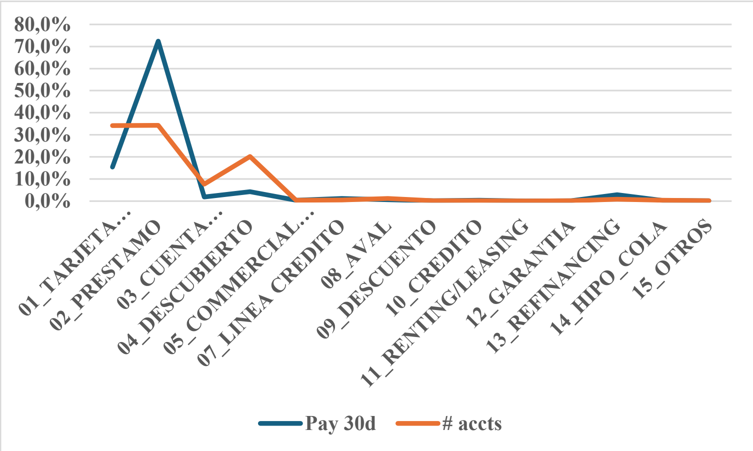
Source: Own elaboration

Discrimination Bivariate Analysis for Individuals and SMEs

In this case, we observe a correlation between the product type grouping and payments in the following 30 days (the distributions of accounts and payments differ across the different product type groupings). For Individuals, the probability for a payment is higher for "Prestamo" – Loans, "Linea crédito" – Credit Lines, and Refinancing. In the case of SMEs, the probability is higher for "Prestamo" – Loans, "Linea crédito" – Credit Lines, "Descuento" – Credit discount and "Otros" – Other. The information value (IV) of this variable is 0.80 for Individuals and 1.05 for SMEs.

Table 43 and Figure 49. Bivariate analysis product type grouped – payment 30 days for Individuals

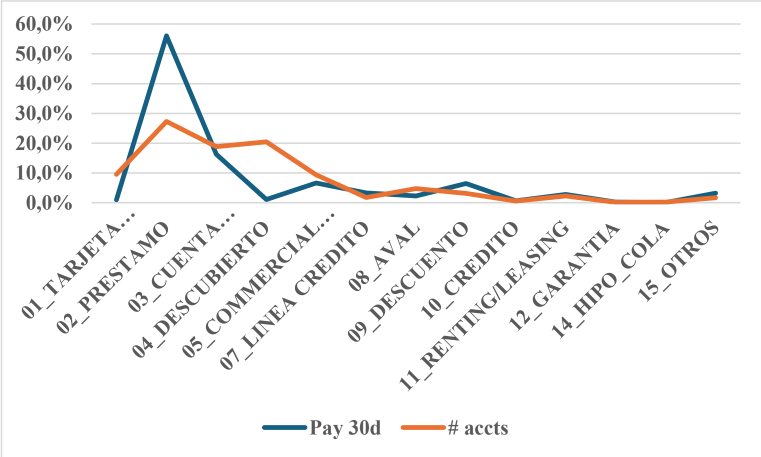
Product type grouped	Pay 30d	# accts
01_TARJETA CREDITO	15,4%	34,2%
02_PRESTAMO	72,4%	34,3%
03_CUENTA CORRIENTE	1,9%	7,6%
04_DESCUBIERTO	4,3%	20,2%
05_COMMERCIAL PAPER	0,4%	0,3%
07_LINEA CREDITO	1,2%	0,5%
08_AVAL	0,6%	1,1%
09_DESCUENTO	0,1%	0,1%
10_CREDITO	0,4%	0,1%
11_RENTING/LEASING	0,0%	0,0%
12_GARANTIA	0,2%	0,1%
13_REFINANCING	2,9%	0,8%
14_HIPO_COLA	0,2%	0,4%
15_OTROS	0,1%	0,2%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 44 and Figure 50. Bivariate analysis product type grouped – payment 30 days for SMEs

Product type grouped	Pay 30d	# accts
01_TARJETA CREDITO	1,0%	9,5%
02_PRESTAMO	56,1%	27,3%
03_CUENTA CORRIENTE	16,2%	18,8%
04_DESCUBIERTO	1,1%	20,5%
05_COMMERCIAL PAPER	6,6%	9,4%
07_LINEA CREDITO	3,3%	1,7%
08_AVAL	2,3%	4,7%
09_DESCUENTO	6,4%	3,2%
10_CREDITO	0,6%	0,5%
11_RENTING/LEASING	2,8%	2,3%
12_GARANTIA	0,2%	0,2%
14_HIPO_COLA	0,1%	0,2%
15_OTROS	3,2%	1,7%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.4.2.10 Variable "IN40_PTP _EVER" – Promise to Pay Ever Flag

Description

This variable reflects whether the account has had a Promise to pay (PTP) for every month of the analysis period. For the construction of this flag, the last available data point for any given month was used. This variable was included in the data set because it typically shows a correlation with the probability of collection. The collection and recovery actions (SMS, calls…) are positively correlated with the probability of collection: the more collection actions, the higher the probability of collection. This variable was not relevant in our modelling in both verticals, Individuals and SMEs.

Individuals' details

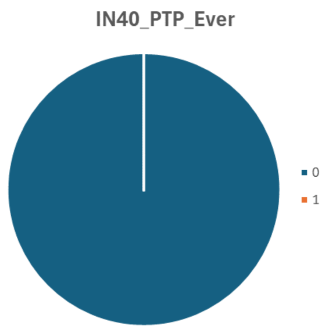
Analysing the last available data point for each bionomy month-account, we observe that only 20 of the total accounts had a promise to pay ever (in the analysis period).

Table 45. Distribution of Individuals accounts by PTP Ever flag

IN40_PTP_Ever	Record_Count
0	676.008
1	20

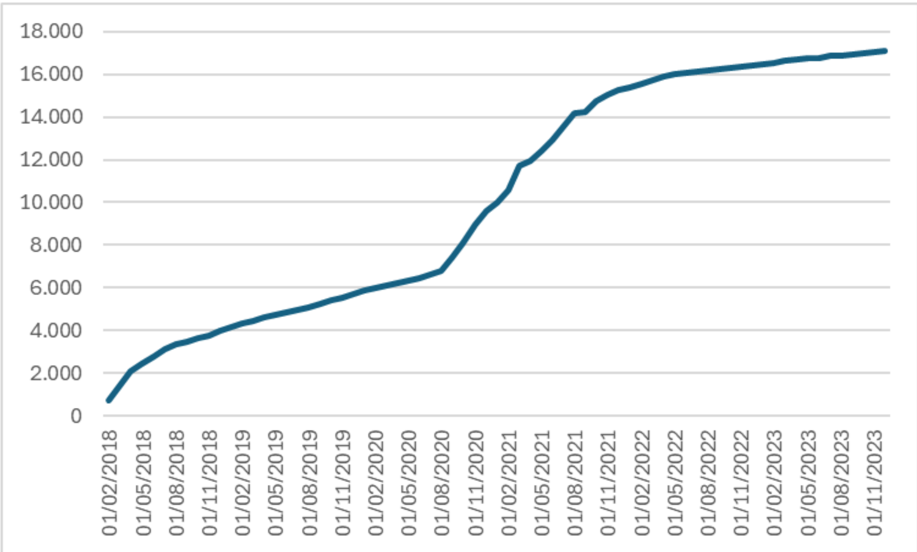
Source: Own elaboration

Figure 51. Distribution of Individuals accounts by PTP Ever flag



Source: Own elaboration

Figure 52. Monthly evolution of Individuals accounts by PTP Ever flag



Source: Own elaboration

SMEs details

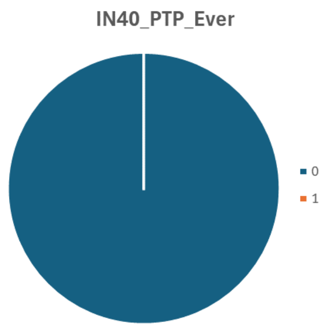
Analysing the last available data point for each bionomy month-account, we observe that none of the accounts had a promise to pay ever (in the analysis period).

Table 46. Distribution of SMEs accounts by PTP Ever flag

IN40_PTP_Ever	Record_Count
0	195.962
1	0

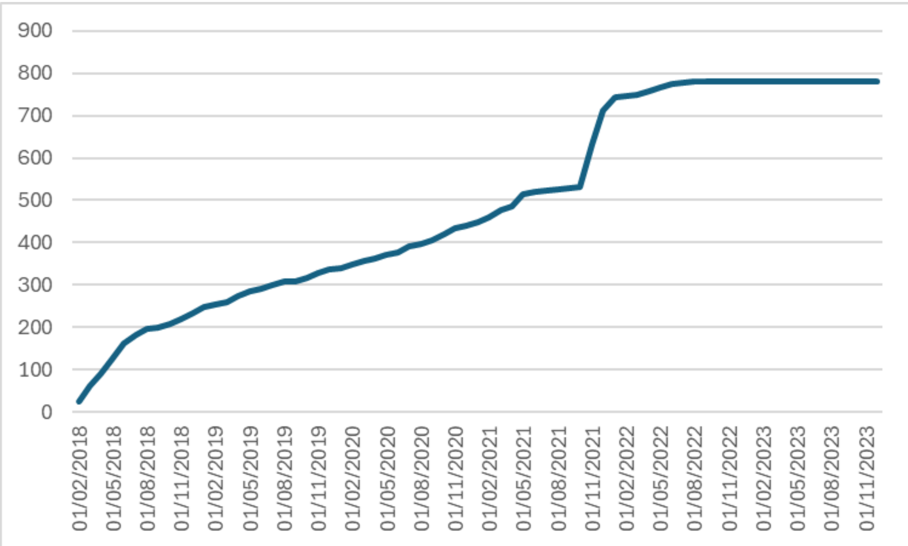
Source: Own elaboration

Figure 53. Distribution of SMEs accounts by PTP Ever flag



Source: Own elaboration

Figure 54. Monthly evolution of SMEs accounts by PTP Ever flag



* No new PTP after sep-22

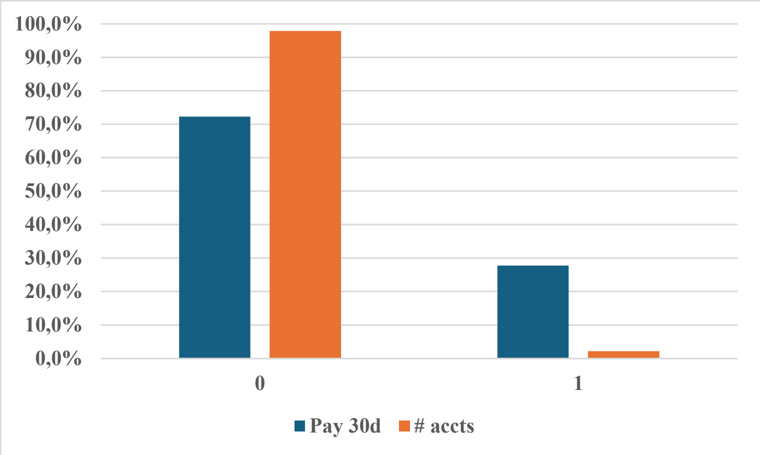
Source: Own elaboration

Discrimination Bivariate Analysis for Individuals and SMEs

In this variable, we observe a correlation between the PTP ever flag and payments in the following 30 days (the distributions of accounts and payments differ between positive and negative flag outcomes). In both Individuals and SMEs, the probability of payment is higher when the flag result is positive. The information value (IV) of this variable is 0.73 for Individuals and 0.07 for SMEs.

Table 47 and Figure 55. Bivariate analysis PTP ever flag – payment 30 days for Individuals

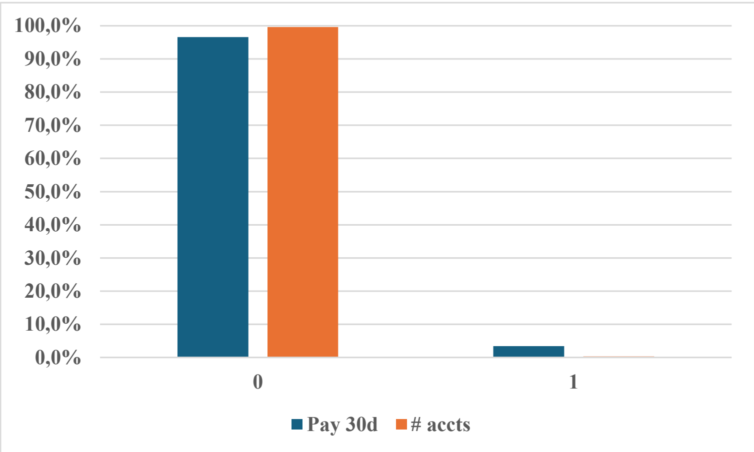
PTP ever flag	Pay 30d	# accts
0	72,2%	97,8%
1	27,8%	2,2%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 48 and Figure 56. Bivariate analysis PTP ever flag – payment 30 days for SMEs

PTP ever flag	Pay 30d	# accts
0	96,6%	99,7%
1	3,4%	0,3%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.4.2.11 Variable "IN22_TP_AVA" – Guarantor Flag

Description

This variable indicates whether the account has a guarantor or a guarantor company.

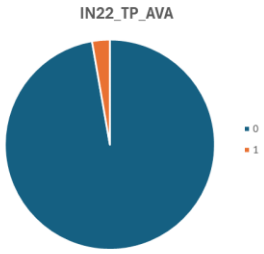
This variable was included in the data set because it typically shows a correlation with the probability of collection. If the account has a guarantor, the probability of collecting is higher than the opposite. This variable was relevant in our modelling only for the SMEs models (within all of them, it is the variable with the highest weight).

Individuals' details

Just 3% of the accounts have a guarantor.

Table 49 and Figure 57. Distribution of Individuals accounts by guarantor flag

IN22_TP_AVA	Record_Count
0	657.495
1	18.533



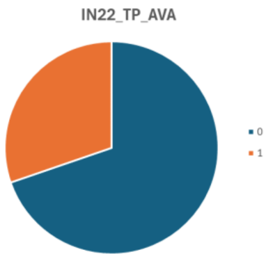
Source: Own elaboration

SMEs details

Around 30% of accounts have a guarantor.

Table 50 and Figure 58. Distribution of SMEs accounts by guarantor flag

IN22_TP_AVA	Record_Count
0	136.620
1	59.342



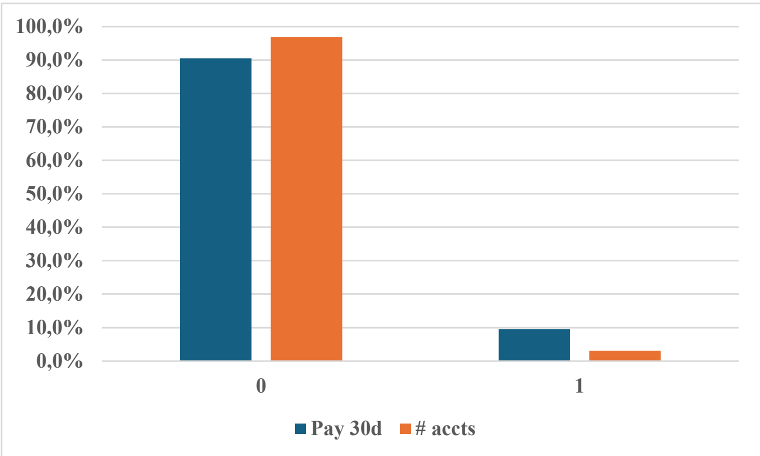
Source: Own elaboration

Discrimination Bivariate Analysis for Individuals and SMEs

In this case, we observe a specific correlation between the guarantor flag and payments in the following 30 days (the distributions of accounts and payments differ across guarantor flags). The probability of a payment is higher when the guarantor flag is positive, and the difference in behaviour is much greater in SMEs than in Individuals. The information value (IV) of this variable is 0.08 for Individuals and 1.04 for SMEs.

Table 51 and Figure 59. Bivariate analysis guarantor flag – payment 30 days for Individuals

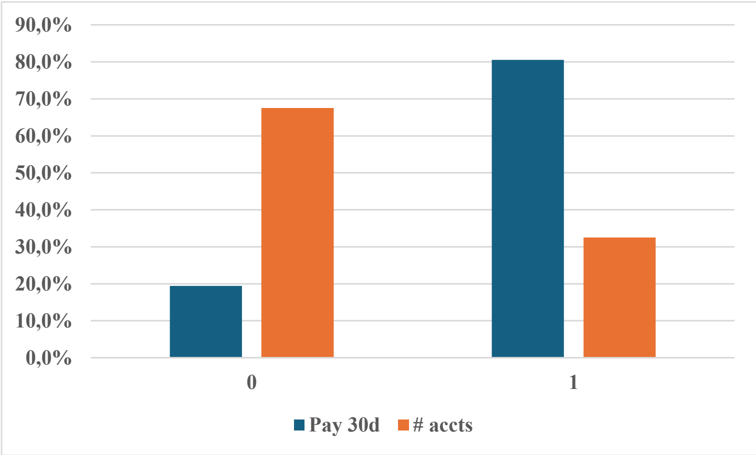
Guarantor flag	Pay 30d	# accts
0	90,5%	96,9%
1	9,5%	3,1%
Total general	100,0%	100,0%



Source: Own elaboration

Table 52 and Figure 60. Bivariate analysis guarantor flag – payment 30 days for SMEs

Guarantor flag	Pay 30d	# accts
0	19,4%	67,5%
1	80,6%	32,5%
Total general	100,0%	100,0%



Source: Own elaboration

3.4.2.12 Variable "IN27_Judicial" – Judicial Flag

Description

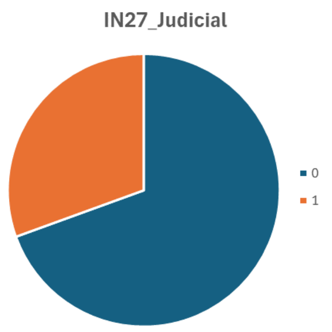
This variable indicates whether the account is or has been subject to a judicial claim after the Investment Firm's acquisition. This variable was included in the data set because it typically shows a correlation with the probability of collection. Accounts with judicial claims are more likely to be collected, as collection efforts are more intensive in these cases. It is also very relevant to note the details of the different statuses or milestones of the judicial claim to understand the predictability of the collection rate. However, we did not have this information available. This variable was relevant for all the models developed within our research.

Individuals' details

14% of the total accounts have been or are currently involved in a judicial claim.

Table 53 and Figure 61. Distribution of Individuals accounts by judicial flag

IN27_Judicial	Record_Count
0	583.373
1	92.655



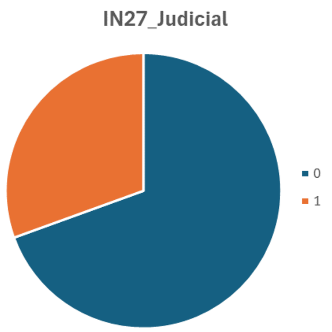
Source: Own elaboration

SMEs details

31% of the total accounts have been or are currently involved in a judicial claim.

Table 54 and Figure 62. Distribution of SMEs accounts by judicial flag

IN27_Judicial	Record_Count
0	136.133
1	59.829



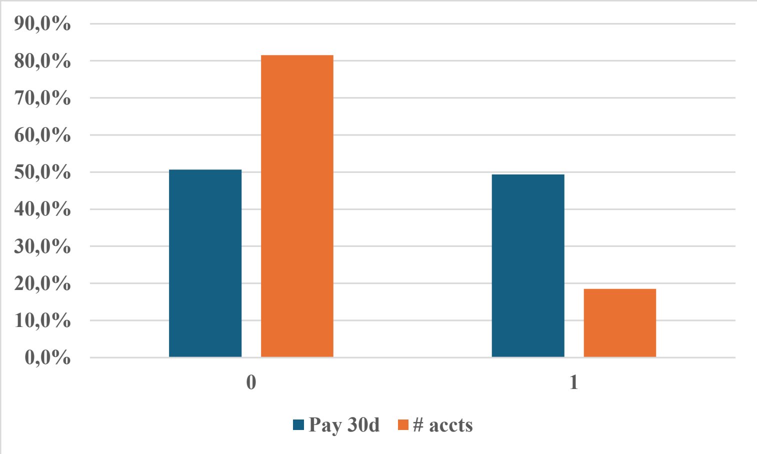
Source: Own elaboration

Discrimination Bivariate Analysis for Individuals and SMEs

In this case, we observe a correlation between the judicial flag and payments in the following 30 days (the distributions of accounts and payments differ across judicial flags). The probability of a payment is higher when the judicial flag result is positive, both for Individuals and SMEs; however, it is more relevant for SMEs. The information value (IV) of this variable is 0.45 for Individuals and 1.02 for SMEs.

Table 55 and Figure 63. Bivariate analysis judicial flag – payment 30 days for Individuals

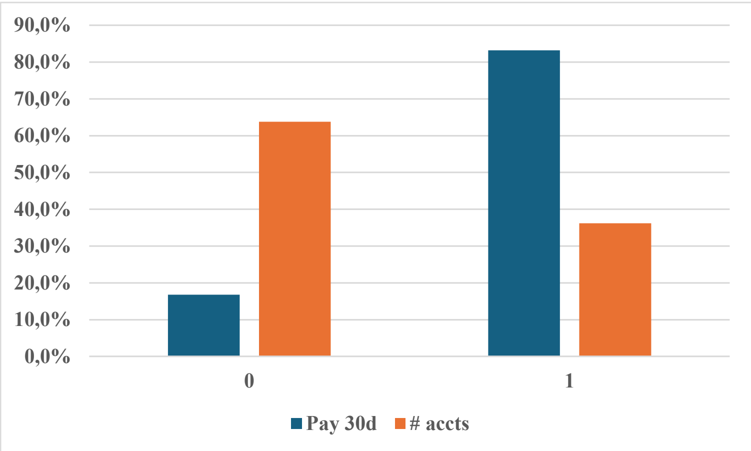
Judicial flag	Pay 30d	# accts
0	50,7%	81,5%
1	49,3%	18,5%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 56 and Figure 64. Bivariate analysis judicial flag – payment 30 days for SMEs

Judicial flag	Pay 30d	# accts
0	16,8%	63,8%
1	83,2%	36,2%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.4.2.13 Macroeconomic Factors

The macroeconomic factors were obtained from the public INE website (https://www.ine.es/), and deciles were constructed for each variable to create a categorical variable from the downloaded data. Table 57 presents the list of macroeconomic variables used in the modelling process, along with each variable's final relevance. The significance of each variable was assessed based on its statistical predictive power. The statistical algorithm used in model development identifies which variables have the highest predictive power, determining their relevance and inclusion in the models.

Table 57. Macroeconomic factors and their relevance for the constructed models

Macroeconomic factors		Models' relevance
"IPC General" – General Consumer Price Index	"IN44_IPC_IndGeneral_VarAnual" – General Consumer Price Index on annual variation	This variable was relevant for the Individuals Global, Individuals COVID, Individuals Post-COVID, SMEs Global, SMEs Pre-COVID, and SMEs Post-COVID models.
	"IN45_IPC_IndGeneral_VarAnual_TIL10" – General Consumer Price Index on annual variation grouped in deciles	This variable was relevant for all the Individuals models, SMEs Global, and SMEs COVID models.
	"IN46_IPC_IndGeneral_VarMensual" – General Consumer Price Index on monthly variation	This variable was relevant for the Individuals Post-COVID, SMEs Pre-COVID, and SMEs Post-COVID models.
	"IN47_IPC_IndGeneral_VarMensual_TIL10" – General Consumer Price Index on monthly variation grouped in deciles	This variable was relevant for all the Individuals models and SMEs Global model.
"IPC Subyacente" – Underlying Consumer Price Index	"IN48_IPC_IndSubyacente_VarAnual" – Underlying Consumer Price Index on annual variation	This variable was relevant for the Individuals Post-COVID and SMEs Post-COVID models.
	"IN49_IPC_IndSubyacente_VarAnual_TIL10" – Underlying Consumer Price Index on annual variation grouped in deciles	This variable was relevant for the Individuals Global, SMEs Pre-COVID, SMEs COVID, and SMEs Post-COVID models.
	"IN50_IPC_IndSubyacente_VarMensual" – Underlying Consumer Price Index on monthly variation	This variable was relevant for the Individuals Pre-COVID and SMEs Post-COVID models.
	"IN51_IPC_IndSubyacente_VarMensual_TIL10" – Underlying Consumer Price Index on monthly variation grouped in deciles	This variable was relevant for the Individuals Global, Individuals Pre-COVID, Individuals COVID, and SMEs COVID models.
"Índice de Confianza del	"IN52_ICC" – Consumer Confidence Index	This variable was relevant for the Individuals Pre-COVID and SMEs COVID models.

Consumidor (ICC)" – Consumer Confidence Index	"IN53_Ind_Sit_Actual" – Consumer Confidence Index actual situation	This variable was relevant for the Individuals Pre-COVID, Individuals COVID, SMEs Global, and SMEs Pre-COVID models.
	"IN54_Ind_Expectativas" – Consumer Confidence Index expectations	This variable was relevant for the Individuals Pre-COVID, Individuals COVID, and SMEs COVID models.
	"IN55_ICC_TIL10" – Consumer Confidence Index grouped in deciles	This variable was relevant only for the SMEs Global model.
	"IN56_Ind_Sit_Actual_TIL10" – Consumer Confidence Index actual situation grouped in deciles	This variable was relevant for the Individuals Global, Individuals Pre-COVID, SMEs Global, SMEs Pre-COVID, and SMEs COVID models.
	"IN57_Ind_Expectativas_TIL10" – Consumer Confidence Index expectations grouped in deciles	This variable was relevant for the Individuals Global, Individuals COVID, SMEs Pre-COVID, and SMEs COVID models.
"Tasa de Actividad" – Activity Rate and "Tasa de Paro" – Unemployment Rate	"IN58_Paro_TasaActividad_VarAnual" – Activity Rate on annual variation	This variable was relevant for the Individuals COVID, Individuals Post-COVID, SMEs Global, and SMEs Post-COVID models.
	"IN59_Paro_TasaActividad_VarAnual_TIL10" – Activity Rate on annual variation grouped in deciles	This variable was relevant for the Individuals Global and Individuals COVID models.
	"IN60_Paro_TasaParo_VarAnual" – Unemployment Rate on annual variation	This variable was relevant only for the SMEs COVID model.
	"IN61_Paro_TasaParo_VarAnual_TIL10" – Unemployment Rate on annual variation grouped in deciles	This variable was relevant only for the SMEs COVID model.
"PIB y Empleo equivalente a tiempo completo" – GDP Full-Time Equivalent Employment	"IN62_PIB_EmpleoEquivTiempoCompleto_VarAnual" – GDP Full-Time Equivalent Employment on annual variation	This variable was relevant for the Individuals Pre-COVID, SMEs Global, and SMEs Pre-COVID models.
	"IN63_PIB_EmpleoEquivTiempoCompleto_VarAnual_TIL10" – GDP Full-Time Equivalent Employment on annual variation grouped in deciles	This variable was relevant for the Individuals Global, Individuals COVID, Individuals Post-COVID, and SMEs COVID models.
	"IN64_PIB_EmpleoEquivTiempoCompleto_VarTrim" – GDP Full-Time Equivalent Employment on quarterly variation	This variable was relevant for the Individuals COVID, SMEs Global, and SMEs COVID models.

	"IN65_PIB_EmpleoEquivTiempoCompleto_VarTrim_TIL10" – GDP Full-Time Equivalent Employment on quarterly variation grouped in deciles	This variable was relevant for the Individuals Pre-COVID, Individuals Post-COVID, and SMEs COVID models.
"PIB a precios de mercado" – GDP Market Prices	"IN66_PIB_PreciosMercado_VarAnual" – GDP Market Prices on annual variation	This variable was relevant for the Individuals Global and SMEs Global models.
	"IN67_PIB_PreciosMercado_VarAnual_TIL10" – GDP Market Prices on annual variation grouped in deciles	This variable was relevant for the Individuals Global, Individuals Pre-COVID, Individuals Post-COVID, and SMEs COVID models.
	"IN68_PIB_PreciosMercado_VarTrim" – GDP Market Prices on quarterly variation	This variable was not relevant for any constructed model.
	"IN69_PIB_PreciosMercado_VarTrim_TIL10" – GDP Market Prices on quarterly variation grouped in deciles	This variable was relevant for the Individuals Global and SMEs COVID models.

Source: Own elaboration

3.4.3 Eliminated Variables and Missing Values Treatment

Upon constructing the data sets and conducting a thorough analysis of the variables, several factors were deemed extraneous for the modelling process due to their limited predictive relevance. Specifically, the variables "IN01_Cartera_g" (portfolio), "IN02_Vertical" (vertical segment), "IN03_Originador" (creditor in its ungrouped form), and "IN04_Tipo_Producto" (type of product in its ungrouped form) were excluded. The portfolio variable was disregarded as the models were designed to encompass a comprehensive array of portfolios, thereby negating the need to recalibrate the models with each new portfolio acquired by the investor. The vertical variable was similarly eliminated, due to the distinct modelling structures established for the various segments, namely Individuals and SMEs, rendering its inclusion redundant.

Regarding the creditor and product type variables, both had excessively high numbers of distinct categories—approximately 1,093 for creditors and 263 for product types. In response, a grouped version of these categories was created to enhance manageability, thereby rendering the ungrouped variants irrelevant for the model.

Moreover, the Method of Payment and SMS variables were also eliminated. The Method of Payment ("METHOD_PAYMENT") was populated solely with values indicating amicable and judicial options, rendering it redundant alongside other variables in the dataset, such as the judicial flag and payments flag, and leading to its elimination. Regarding the "SMS" variables (namely, "IN29_SMS_L30D," "IN33_SMS_L90D," and "IN37_SMS_EVER"), they were excluded due to their excessive correlation with the target predictive variable. It was determined that the high correlation did not indicate that SMS communications catalysed payments; rather, it stemmed from a non-random selection process for accounts designated to receive SMS notifications. This observation was corroborated through a comparative analysis with the Investment Firm, resulting in consensus on the interpretation.

When addressing missing values or incomplete data, the initial step is to identify the cause of the missing data and evaluate how significantly it affects the overall dataset. It is crucial to ascertain whether the data loss is random (Missing Completely at Random - MCAR), dependent on other variables (Missing at Random - MAR), or influenced by the missing values themselves (Missing Not at Random - MNAR). If only 1-5% of the data is absent, the choice of method becomes less critical. However, if over 20-30% is missing for a particular variable, it may be more prudent to discard that variable or to employ advanced imputation techniques.

The data obtained from the Investment Firm shows that there is not a significant amount of missing values. However, we can categorize the affected variables into two types. First, there are variables with a limited number of missing entries, such as Originator or Province.

In these cases, a single field has been created to capture all the erroneous or empty values.

Second, certain variables with numerous affected fields, specifically the Judicial Claim Date ("FECHA_DEMANDA") and Judicial Procedure Status ("ESTADO_PROC"), exhibited a substantial proportion of unreported fields—exceeding 38%. Additionally, a significant IT systems migration in 2020 led to approximately 53% of reported entries capturing the migration date instead of the actual Judicial Claim Date, thereby undermining the integrity of these fields for model construction. In such instances, it has been deemed preferable to utilize highly correlated variables that provide more comprehensive information, such as "Judicial," which indicates whether the file is currently or has been involved in a judicial situation.

SECTION 4.- MODELLING METHODOLOGY AND RESULTS

4.1.- MODELLING FRAMEWORK

4.2.- METHODOLOGY USED

4.3.- CONSTRUCTED MODELS

4.1.- MODELLING FRAMEWORK

To achieve the research objective of identifying the debt portfolio segmentation variables that most effectively predict collection rates in written-off portfolios, and to evaluate the impact of COVID-19 (and all measures implemented to alleviate the pandemic effects) on their significance (which variables significantly predict short-term repayment behaviour before, during, and after the COVID-19 crisis?), we used the available data for our analysis. This data includes written-off portfolios acquired by the Investment Firm from 2015 to mid-2022, alongside their management and payment behaviour observed between 2018 and 2023. We developed statistical models to estimate the likelihood that an account will make payments in the subsequent months under the Investment Firm's management.

The models were developed to differentiate between Individual and SME debtors, acknowledging their distinct characteristics and payment behaviours. To analyse the variation in the importance of determinants influenced by the impact of COVID-19, several models were created: one model encompasses the entire period from 2018 to 2023 (“Global”); a second model covers the timeframe preceding the pandemic, specifically 2018-2019 (“Pre-COVID”); the third model addresses the period during the pandemic, which includes 2020-2021 (“COVID”); and, lastly, a model is dedicated to the timeframe following the pandemic, covering 2022-2023 (“Post-COVID”). In total, eight models were constructed, as detailed in Table 58.

Table 58. Summary of constructed models

Segment	Models
Individuals	• Individuals Global model • Individuals Pre-COVID model • Individuals COVID model • Individuals Post-COVID model
SMEs	• SMEs Global model • SMEs Pre-COVID model • SMEs COVID model

	• SMEs Post-COVID model

Source: Own elaboration

Once the models were developed, we analysed the variables included in each model and assessed their importance weight. The final number of variables ranged from 19 to 26, as follows: Individuals Global – 25 variables; Individuals Pre-COVID – 26; Individuals COVID – 24; Individuals Post-COVID – 22; SMEs Global – 21; SMEs Pre-COVID – 19; SMEs COVID – 24; and SMEs Post-COVID – 19. However, the significance of these variables was primarily concentrated on 9 to 12 of them, with the remaining variables offering marginal value in enhancing the models' predictive capabilities.

Given this context, we opted to create two versions of the models. The first version includes all the variables, thereby maximizing their predictive power and providing greater practical utility ("complete" models). The second version focuses solely on the most relevant variables, which, while offering lower predictive power, are easier to explain and interpret. This version is expected to be more beneficial from an academic standpoint ("reduced" models).

4.2.- METHODOLOGY USED

4.2.1.- Training and Testing Samples, and Overfitting Treatment

Once the data sets were assembled with the input variables and the target variable, the first step was to split the samples for modelling: one sample for model construction and training, and another for testing to ensure the robustness and accuracy of the developed model, thereby avoiding overfitting.

Overfitting is a common challenge encountered in the development of predictive models. It arises primarily when the model is excessively complex relative to the quantity or quality of the available training data. As a result, the model may learn not only valuable underlying patterns but also random noise and irrelevant features present in the training set.

Throughout the model construction process, both training and test samples were consistently included, with the test samples maintained separately from the modelling perimeter to prevent any interference with the generated rules. Once the final model was completed, it was evaluated using the parallel test sample to assess its performance. All tests conducted for the Pre-COVID, COVID, and Post-COVID models revealed remarkably similar results between the models developed from the training samples and those tested on the test samples.

The sample selections were entirely independent and random, made at the outset of the modelling process, and at the account level (not at the observation level), to ensure there is no contamination of the testing sample. A typical division was used, allocating 80% of the accounts of the data sets to the training sample and 20% to the testing sample (Li et al., 2022). This random selection was performed using a seed in IBM SPSS Modeler, the tool used to manage the data and construct the statistical models.

Furthermore, once the model is deployed, the newly generated samples (representing the new months and all their associated variables) can serve as an additional validation dataset. This will help ensure that there is no overfitting in the developed model and act as a control measure to prevent model miscalibration.

4.2.2.- Modelling Methodology

As delineated in the Literature Review section of this document, there exists a notable lack of consensus within the academic community regarding the most effective methodology for estimating probabilities in data collection. Nevertheless, the predominant trends in methodological approaches can be categorized into three principal types: traditional regression techniques, classification and regression trees (CART), and advanced machine learning methodologies, which encompass deep learning, random forests, and boosting algorithms.

The foundational work of Breiman and Friedman in 1973 introduced recursive partitioning algorithms, with classification and regression trees (CART) emerging as the most frequently employed tree algorithms, demonstrating robust performance across various applications (Malikkidou & Strohbach, 2025; Xu et al., 2024; Thomas & Bijak, 2015; Bijak, 2014). Notably, classification trees, in conjunction with random forests, have frequently achieved superior accuracy compared with alternative methods (Malikkidou & Strohbach, 2025; Li et al., 2024).

Regression trees represent a potent yet straightforward technique for regression analysis, wherein the predicted values of the target variable are derived through a series of sequential logical if-then conditions. This methodical sequence of binary splits effectively partitions observations based on loan characteristics. The algorithm initiates with a root node encompassing all observations, subsequently searching through all binary splits of available attributes to identify the split that minimizes the intra-subset variation of the target variable in the resulting daughter node. Consequently, each daughter node demonstrates a greater homogeneity of the target variable than its parent node. This recursive procedure continues until no further reduction in the variation of the target variable is attainable (Bastos, 2014; Bastos, 2010).

For the present research, the chosen methodological framework is a CHAID (Chi-square Automatic Interaction Detector) classification tree. This methodology employs a decision tree algorithm primarily intended for classification and segmentation tasks, and it is notably characterized by its approach to tree construction, which accommodates non-binary divisions. Nodes in the tree are divided into two or more branches, in contrast to binary methods. Within this framework, the dependent variable must be categorical (either nominal or ordinal), while the predictor independent variables with fewer cases are combined to ensure that the resulting daughter nodes meet a minimum required size.

For each node, CHAID assesses every available predictor variable. It starts by merging the categories of each predictor that do not significantly differ concerning the dependent variable. Then, it evaluates all possible splits of the predictor (with or without merged categories) and uses the Chi-square test or the likelihood ratio to identify the predictor that has the most significant relationship with the dependent variable at that particular node. The variable exhibiting the highest Chi-square value, along with a p-value below a predetermined threshold (defaulting to 0.05), is chosen to split the node. This process is repeated until no predictors are deemed significant or until a stopping criterion, such as minimum node size, is met.

This approach is distinguished by its robustness in the presence of outlier data, a common occurrence in collection activities. Additionally, the stability of the tree structure is maintained even as individual variables within the models evolve over time. The clarity and interpretability of these models, due to their transparency regarding determinants and their respective weights and importance, render them accessible. Furthermore, the implementation and maintenance of CHAID models are comparatively straightforward, especially when juxtaposed with more complex methodologies such as XGBoost or random forests. Beyond its academic recognition, CHAID is widely employed in business practices, particularly within the Analytics Departments of institutions engaged in collections and recovery activities, including banks, investment funds, and debt collection agencies.

The general specifications for the model's construction utilized in this research are detailed in Table 59.

Table 59. General specifications for models construction

Tree levels used	7
Alpha for division	0.05
Alpha for melting	0.05
Epsilon for convergence	0.001
Maximum number of iterations for convergence	100
Using Bonferroni's correction	Yes
Chi-square method	Pearson
Minimum Parent Branch number of Records	2%
Minimum Child Branch number of Records	1%

Source: Own elaboration

In line with emerging trends in literature regarding the application of machine learning techniques, we explored various algorithmic methodologies for this research, including Random Forest and XGBoost (Malikkidou & Strohbach, 2025; Xu et al., 2024). We evaluated these methodologies against the CHAID classification tree, focusing on their predictive power and ease of implementation, ultimately deciding to exclude them from further consideration. The primary distinction of the CHAID tree, in comparison to other ensemble methods such as Random Forest or Boosting, lies in its capacity to elucidate interactions among multiple predictors in a transparent and visually interpretable manner. This characteristic facilitates a deeper understanding of the relationships within the data, allowing for the identification of complex interactions that may not be readily apparent in more opaque models.

The tests indicated that the Random Forest model did not yield any improvements over the CHAID tree. This may be attributed to several factors, such as the CHAID tree already being effectively optimized and pruned, resulting in a favourable balance between bias and variance. Alternatively, it is possible that the hyperparameters of the Random Forest did not enable sufficient decorrelation among the various trees.

In the comparison of CHAID and XGBoost, it was noted that the Boosting model demonstrated superior performance indicators. However, the enhancement was minimal, with less than a one-point increase in the Gini coefficient. Consequently, the decision was made to proceed with the CHAID tree for several reasons: its simplicity in model development, time efficiencies in execution, ease of implementation in a production environment, straightforward maintenance, and clarity in analysing results and identifying the most relevant variables.

4.2.3.- Predictability Estimations

In any statistical modelling construction process, it is essential to evaluate the robustness, confidence, and sensitivity of the developed models. In this case, we need to assess the accuracy of the predictions generated by our models, which estimate the likelihood of an account making a payment within the next 30 days.

To conduct this assessment, we have opted to employ some of the established and classical methods in classification tree modelling (Malikkidou & Strohbach, 2025; Li et al., 2023B; Dumitrescu et al., 2022; Hoechstoetter et al., 2012): to evaluate the performance and predictive power of the models, we will utilize the Receiver Operating Characteristics curve (ROC curve), the Area Under the Curve analysis (AUC), and the Gini coefficient;

and, additionally, to ensure the representativeness of the data used and the models developed, we have implemented the Kolmogorov-Smirnov test (KS test).

The Receiver Operating Characteristic (ROC) curve is a graphical representation of the trade-off between the true positive rate and the false positive rate across various classification thresholds (Malikkidou & Strohbach, 2025). This methodology is widely regarded as a robust approach to evaluating the performance of predictive models at different cut-off points, yielding a metric known as the Area Under the Curve (AUC) (Li et al., 2023B). The AUC quantitatively assesses a model's or classifier's overall discriminatory performance (Dumitrescu et al., 2022). An AUC value approaching 1 signifies optimal classification capability, whereas an AUC of 0.5 indicates a lack of discriminative power (Li et al., 2023B).

Moreover, the Gini coefficient, which is equivalent to double the area between the ROC curve and the diagonal line, acts as a direct transformation of the AUC. It evaluates the discriminative effectiveness of classifiers across multiple thresholds, with values approaching 1 indicating perfect classification performance (Dumitrescu et al., 2022; Lesssmann et al., 2015).

The Gini coefficient is commonly used in classification trees (Nembrini et al., 2018), and it has a long academic history, as more than 100 years ago, in 1912, Corrado Gini presented the initial version of this best-known measure of concentration based on the Lorenz curve of income distribution (Laitinen, 2025). As Laitinen mentions, the Gini coefficient, along with AUC, is a very popular measure for evaluating the accuracy of predictive models in credit rating. The question is how these concentration metrics assess the predictive model's results, where “bad events” (the desired outcome) and “good events” are ranked. The goal is that bad events should concentrate at the top of the order as strongly as possible. Then, the Gini coefficient is used to assess the distribution of bad events in the rank order. The more bad events are concentrated at the front of the ranking produced by the predictive model, the better the model's performance is considered by this metric.

As an example applied to our models, which predict the likelihood of an account making a payment in the next 30 days: if the model has a Gini of 0.55, it means that approximately 20% of accounts top-scored might deliver 50–60% of all payments, and the bottom 20% might deliver almost none.

In C&R activities and modelling, Gini values below 0.40 are considered weak, 0.40-0.60 are good, and above 0.60 are excellent.

The Gini coefficient and AUC indicator are highly related, and we decided to use the Gini coefficient as the primary reference for our models' assessment due to its clearer, more intuitive interpretation, its use in the academic literature (for comparison of results), and its broader use in C&R practice.

The Kolmogorov-Smirnov test, a nonparametric statistical test, is used to determine whether the distributions of two distinct data sets align around their means. This test evaluates the maximum distance between the estimated cumulative distribution functions of two random variables (Dumitrescu et al., 2022).

In general, empirical evidence suggests that the predictability of LGD and RR models is markedly inferior, exhibiting considerably lower performance levels in comparison to Probability of Default (PD) models (Li et al., 2023A; Loterman et al., 2012; Höcht & Zagst, 2010; Dermine & De Carvalho, 2006).

A notable contribution to the field of credit scoring methodologies pertinent to LGD and RR modelling is the research conducted by Dumitrescu et al. They rigorously evaluated various techniques—including Linear Logistic Regression, Non-Linear Logistic Regression, a combination of Non-Linear Logistic Regression and ALasso, Random Forest (RF), Penalised Logistic Tree Regression (PLTR), Support Vector Machine (SVM), and Neural Network (NN)—across four credit default datasets. Table 60 presents a compilation of the Gini and Kolmogorov-Smirnov indicators for the models analysed. The PLTR methodology emerged as the most effective in their study, achieving a Gini coefficient of 0.7076 and a K-S statistic of 0.5647 (Dumitrescu et al., 2022).

Table 60. Gini and K-S indicators of the Dumitrescu models

Models	Gini results	K-S results
Linear Logistic Regression	0.3964	0.3168
Non-Linear Logistic Regression	0.5255	0.4173
Non-Linear Logistic Regression + ALasso	0.6102	0.4751
RF	0.6990	0.5563
PLTR	0.7076	0.5647
SVM	0.4830	0.3723
NN	0.5006	0.3895

Source: Own elaboration

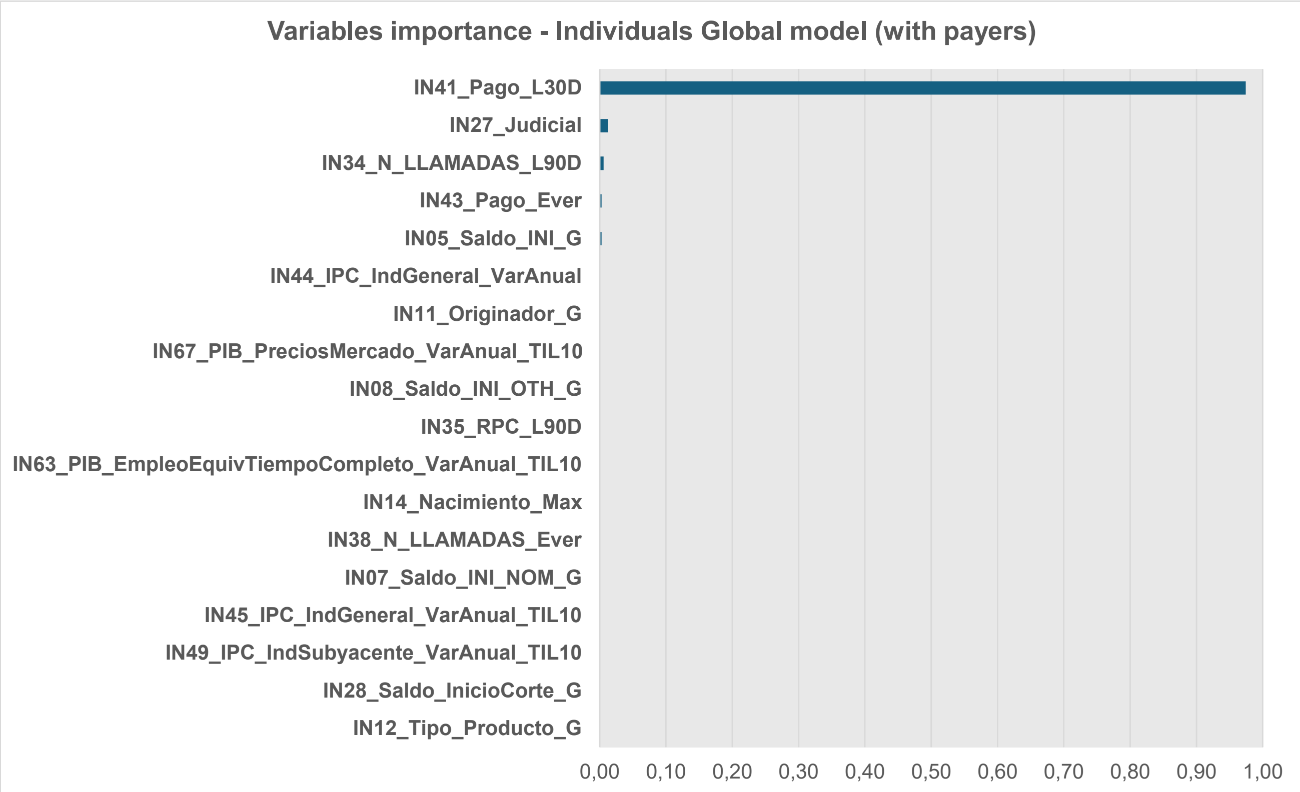
Another example can be found in the research conducted by Bijak & Thomas, which analysed personal loan data from a major UK bank spanning the years 1987 to 1998. This study focused on loans that defaulted between 1988 and 1999, with recovery data extending until 2004. The models developed in their research achieved a Gini coefficient of 0.42 and a K-S statistic of 0.31 (Bijak & Thomas, 2015).

4.2.4.- Payers vs Non-payers Modelling

LGD and RR modelling exhibit a bimodal distribution of recovery rates, with a significant concentration of results at both ends of the spectrum (Calabrese & Zanin, 2022; Ye & Bellotti, 2019; Calabrese & Zenga, 2010; Schuermann, 2004). During model construction, this bimodal distribution was observed, particularly the strong correlation between the payer variable (clients who have previously made partial payments on their outstanding debts) and the target variable (the probability of an account making a payment within the next 30 days).

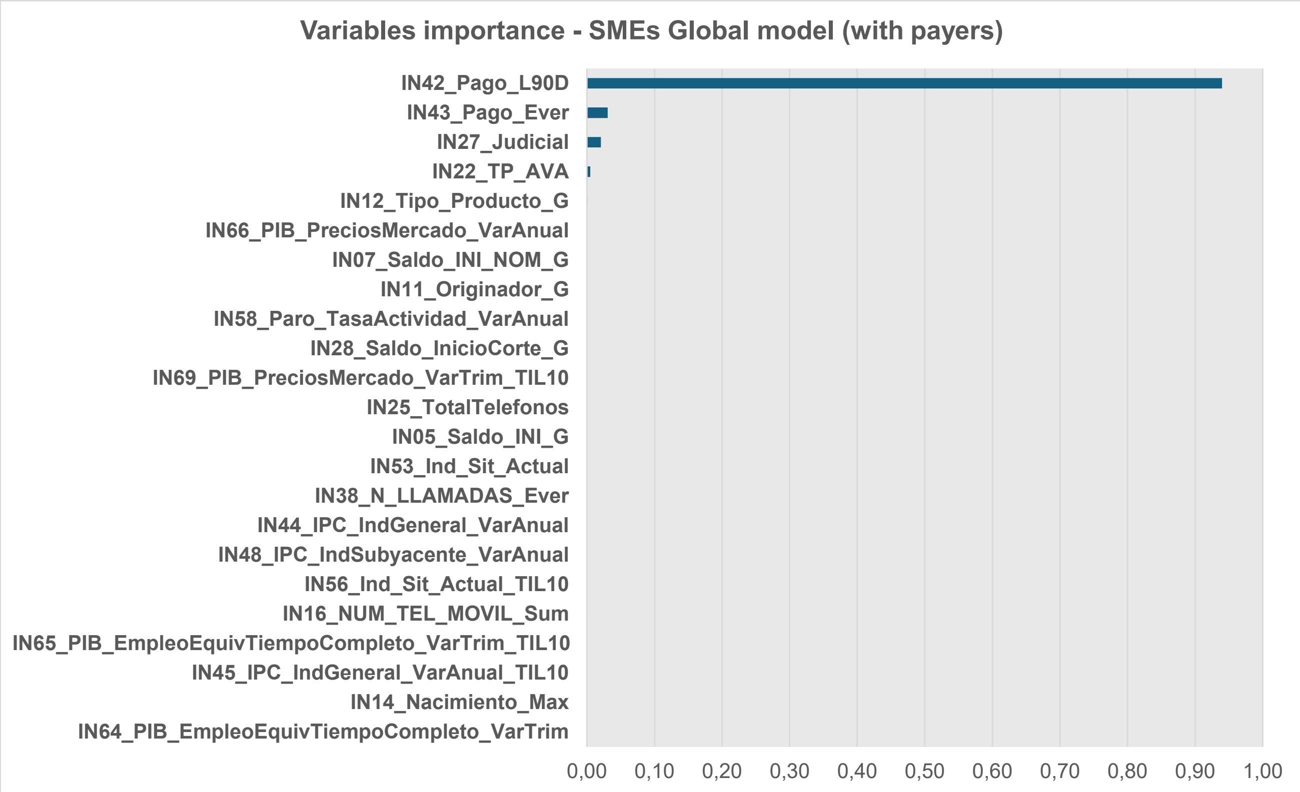
As shown in the subsequent figures and tables, which present the variables alongside their importance and the predictive and robustness statistical indicators for the Individuals Global and SMEs Global models, the models that incorporated payer information demonstrated strong predictive and robustness indicators. However, they were heavily skewed toward the payer variable, significantly diminishing their practical utility.

Figure 65. Variables importance for the Individuals Global model (with payers)



Source: Own elaboration

Figure 66. Variables importance for the SMEs Global model (with payers)



Source: Own elaboration

Table 61. Gini and K-S indicators of the Individuals Global model (with payers)

Indicators	Training sample	Testing sample
Gini	0,903	0,904
K-S	0,787	0,787

Source: Own elaboration

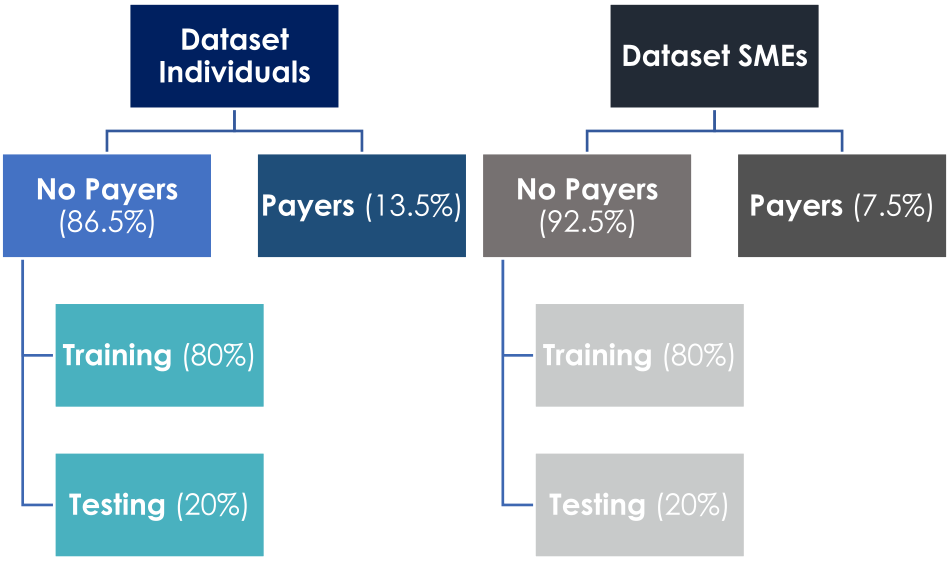
Table 62. Gini and K-S indicators of the SMEs Global model (with payers)

Indicators	Training sample	Testing sample
Gini	0,905	0,900
K-S	0,777	0,774

Source: Own elaboration

Additional to these results identified, it must be remarked that the number of payers accounts is quite reduced within our datasets, in line with the debt age of the accounts considered in our research: as represented in Figure 67, in the Individuals dataset, payers account for 13.5% of the total sample size, whereas in the SMEs dataset, they constitute 7.5% of the total sample size.

Figure 67. Individuals and SMEs data sets split distribution



Source: Own elaboration

Based on the highlighted results from the Global models (which included payer accounts) and the reduced percentage of payer accounts in our datasets, we decided to continue and focus our research by building new versions of the models without considering payer accounts, which are defined as accounts that have previously made payments. In Section 4.3 of the present document, we detailed the models constructed without considering the payers, including a version with all available variables ("completed" models) and a version with only the more relevant variables ("reduced" models).

Despite this, the appendices of this document provide detailed information on the models constructed using payer data.

4.3.- CONSTRUCTED MODELS

In this section, we provide a comprehensive examination of the models developed for the Individuals and SMEs verticals, emphasizing the variables incorporated, their significance within the models, and the associated predictive indicators, specifically the Gini coefficient and Kolmogorov-Smirnov (K-S) statistic.

Initially, the models were constructed using all available variables, referred to as the “complete” models. Analysing these complete models reveals that the number of statistically significant variables included ranged from 19 to 26. Specifically, the Individuals Global model comprised 25 variables, the Individuals Pre-COVID model 26, the Individuals COVID model 24, and the Individuals Post-COVID model 22. For the SME models, the Global model featured 21 variables, the Pre-COVID model 19, the COVID model 24, and the Post-COVID model 19. It is noteworthy that a substantial proportion of these variables exhibited marginal statistical significance.

As a result of these findings, we refined our models by concentrating solely on the most relevant variables, leading to the development of what we refer to as the “reduced” models. This refinement process was performed without sacrificing the models' essential predictive capabilities. The number of variables in these reduced models varied from 7 to 11: the Individuals Global model incorporated 9 variables, the Individuals Pre-COVID model included 8, the Individuals COVID model contained 7, and the Individuals Post-COVID model featured 10 variables. For SMEs, the Global model also included 9 variables, while the Pre-COVID and COVID models featured 11, and the Post-COVID model consisted of 10.

In the subsequent subsections of this document, we present a summary of the variables included in both the complete and reduced models, along with the Gini and K-S indicators for each constructed model. Additionally, detailed information regarding the reduced models will be provided. The details of the complete models are in Appendix 5 of this document.

4.3.1.- Summary of Variables Included in the Models and Their Relevance

In Table 63, we summarize the variables included in the complete models for Individuals, along with their respective weights. The weight assigned to each variable quantifies the extent to which that variable contributes to the outcome predicted by the model. This weight reflects the variable's importance within the constructed framework. The cumulative weight of all the variables included in the model totals 1. Variables with weights approaching 0 suggest they may be dispensable, while those with weights higher than 0.15 to 0.20 are deemed highly relevant to the model.

These variable weights have been analysed for each model built and are detailed in the following sections of this document.

As noted in the literature, the key determinants for the Individuals' models consist of a combination of debt characteristics (such as creditor, debt amount, and product type), collection activities (including judicial flags, RPC, and call attempts), macroeconomic factors (like GDP, GCPI, and AR), and debtor characteristics (such as nationality, date of birth, and the number of phone numbers). Two variables that consistently emerge as highly relevant across all models are the judicial flag and the creditor. Additionally, we observe that many variables exhibit marginal weights in the models (below 0.0050). Consequently, we constructed models focused solely on the most significant variables, as detailed in Table 64, which summarize the variables from the reduced models for Individuals.

Table 63. Model variables and their weight for Individuals complete models

Global		Pre-COVID		COVID		Post-COVID	
Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight
Judicial flag	0.3436	**Judicial flag**	0.3778	**RPC I90d**	0.3713	**Creditor grouped**	0.5822
Creditor grouped	0.2829	**Initial other debt amount**	0.3179	**Judicial flag**	0.3251	**Judicial flag**	0.1712
RPC I90d	0.2387	**Creditor grouped**	0.1875	**Creditor grouped**	0.2817	**RPC I90d**	0.1680
Initial other debt amount	0.0694	**National flag**	0.0531	**Calls ever flag**	0.0056	**Other nationality flag**	0.0243
GDP FTEE annual deciles	0.0180	**Youngest DOB**	0.0176	**RPC ever flag**	0.0042	**Product type grouped**	0.0198
GCPI monthly deciles	0.0116	**Region**	0.0170	**UCPI monthly deciles**	0.0040	**Calls I90d flag**	0.0156
GCPI annual deciles	0.0063	**Initial interest debt amount**	0.0094	**Monthly debt amount grouped**	0.0023	**GCPI annual**	0.0043
National nationality flag	0.0062	**GCPI monthly deciles**	0.0039	**GCPI annual**	0.0017	**Calls ever flag**	0.0028
AR annual deciles	0.0059	**# fixed line phone numbers**	0.0026	**# mobile phone numbers**	0.0016	**UCPI annual**	0.0024
GDP MP quarter deciles	0.0044	**UCPI monthly deciles**	0.0024	**National nationality flag**	0.0011	**RPC ever flag**	0.0024
UCPI monthly deciles	0.0042	**Initial principal debt amount**	0.0024	**Region**	0.0009	**Monthly debt amount grouped**	0.0017
Product type grouped	0.0040	**Calls ever flag**	0.0021	**GDP FTEE quarter**	0.0003	**GDP FTEE annual deciles**	0.0015
Region	0.0028	**GDP FTEE quarter deciles**	0.0021	**AR annual deciles**	0.0001	**GDP MP annual deciles**	0.0013
# phone numbers	0.0017	**Foreign nationality flag**	0.0016	**CCI actual**	0.0001	**GCPI monthly deciles**	0.0012
Calls I90d flag	0.0003	**GCPI annual deciles**	0.0014	**Initial principal debt amount**	0.0000	**GCPI annual deciles**	0.0011
# fixed line phone numbers	0.0000	**UCPI monthly**	0.0008	**GCPI annual deciles**	0.0000	**Calls I30d flag**	0.0002
CCI actual deciles	0.0000	**CCI actual**	0.0003	**Product type grouped**	0.0000	**Initial principal debt amount**	0.0000

CCI expectations deciles	0.0000	**Initial other debt amount**	0.0000	**CCI expectations deciles**	0.0000	**# phone numbers**	0.0000
GDP MP annual	0.0000	**CCI actual deciles**	0.0000	**Initial interest debt amount**	0.0000	**GCPI monthly**	0.0000
Initial interest debt amount	0.0000	**# participants**	0.0000	**GCPI monthly deciles**	0.0000	**GDP FTEE quarter deciles**	0.0000
Initial principal debt amount	0.0000	**GDP MP annual deciles**	0.0000	**GDP FTEE annual deciles**	0.0000	**UCPI monthly**	0.0000
GDP MP annual deciles	0.0000	**CCI**	0.0000	**AR annual**	0.0000	**AR annual**	0.0000
GCPI annual	0.0000	**CCI expectations**	0.0000	**CCI expectations**	0.0000		
UCPI annual deciles	0.0000	**# phone numbers**	0.0000	**Calls l90d flag**	0.0000		
Calls ever flag	0.0000	**Product type grouped**	0.0000				
		GDP FTEE annual	0.0000				

Source: Own elaboration

In the analysis of the Individuals reduced models, it is evident that debtor characteristics exert a minimal influence and presence, with notably no representation in the COVID model. Instead, the predominant factors affecting these models are rooted in debt characteristics, collection activities, and broader macroeconomic elements. Among these, the variables Judicial flag and Creditor emerge as the most significant determinants across most models, exhibiting mixed weightings ranging from 0.6432 to 0.7725. Table 64 provides a comprehensive overview of the variables incorporated in the Individuals reduced models, detailing the weight of each variable within the constructed models.

Table 64. Model variables and their weight for Individuals reduced models

Global		Pre-COVID		COVID		Post-COVID	
Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight
Judicial flag	0.5570	**Judicial flag**	0.4752	**Creditor grouped**	0.3230	**Creditor grouped**	0.5901
RPC l90d	0.2240	**Initial other debt amount**	0.2476	**Judicial flag**	0.3202	**Judicial flag**	0.1824
Creditor grouped	0.1336	**Creditor grouped**	0.1709	**RPC l90d**	0.3143	**RPC l90d**	0.1636

Initial other debt amount	0.0693	**National nationality flag**	0.0514	**Monthly debt amount grouped**	0.0239	**Product type grouped**	0.0201
GDP FTEE annual deciles	0.0116	**Youngest DOB**	0.0314	**Calls ever flag**	0.0097	**Other nationality flag**	0.0192
National nationality flag	0.0034	**Region**	0.0122	**RPC ever flag**	0.0050	**Calls l90d flag**	0.0140
GCPI annual deciles	0.0011	**GCPI monthly deciles**	0.0075	**UCPI monthly deciles**	0.0039	**UCPI annual**	0.0054
GCPI monthly deciles	0.0000	**Initial interest debt amount**	0.0039			**RPC ever flag**	0.0032
AR annual deciles	0.0000					**GCPI annual**	0.0020
						Calls ever flag	0.0000

Source: Own elaboration

In Table 65, we present a summary of the variables included in the complete models for SMEs, along with their respective weights. Similar to the Individual models, and consistent with findings in the literature, the key determinants in the Individual models comprise a blend of debt characteristics (such as debt amount, type of creditor, or product type), collection activities (including judicial flags, RPC, or phone calls), macroeconomic factors (like GDP, GCPI, or CCI), and debtor characteristics (such as the guarantor flag, number of phone contacts, or the date of company establishment). Notably, two variables stand out as the most significant across all constructed models: the Guarantor flag and the Judicial flag. The combined weight of these two variables falls within the range of 0.7277 to 0.8648.

As with the Individual models, we find that many variables exhibit a marginal weight in the SMEs models (below 0.0050). Consequently, we have developed models that incorporate only the most relevant variables. Table 66 displays the variables used in the reduced models for SMEs.

Table 65. Model variables and their weight for SMEs complete models

Global		Pre-COVID		COVID		Post-COVID	
Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight
Guarantor flag	0.4969	**Guarantor flag**	0.4582	**Guarantor flag**	0.4625	**Guarantor flag**	0.5144
Judicial flag	0.3679	**Judicial flag**	0.2695	**Judicial flag**	0.3214	**Judicial flag**	0.2344
Product type grouped	0.0474	**Monthly debt amount grouped**	0.0976	**Initial principal debt amount**	0.0941	**RPC ever flag**	0.1430
Creditor grouped	0.0203	**Product type grouped**	0.0691	**Creditor grouped**	0.0328	**GCPI monthly**	0.0224
Initial interest debt amount	0.0181	**# phone numbers**	0.0316	**CCI**	0.0152	**UCPI annual**	0.0211
Monthly debt amount grouped	0.0150	**Creditor grouped**	0.0162	**UR annual**	0.0122	**Product type grouped**	0.0164
GDP MP annual	0.0070	**Region**	0.0139	**GDP MP quarter deciles**	0.0119	**Creditor grouped**	0.0130
# phone numbers	0.0067	**GCPI monthly**	0.0124	**Initial interest debt amount**	0.0112	**# fixed line phone numbers**	0.0099
# mobile phone numbers	0.0065	**Calls ever flag**	0.0091	**CCI expectations deciles**	0.0094	**GCPI annual**	0.0068
Initial principal debt amount	0.0037	**CCI expectations deciles**	0.0084	**CCI actual deciles**	0.0072	**Monthly debt amount grouped**	0.0062
GDP FTEE annual	0.0034	**# mobile phone numbers**	0.0064	**GDP MP annual deciles**	0.0064	**AR annual**	0.0044
GCPI annual	0.0028	**CCI actual**	0.0056	**Province**	0.0039	**Youngest date of company creation**	0.0041
GCPI monthly deciles	0.0018	**Youngest date of company creation**	0.0009	**GDP FTEE quarter deciles**	0.0037	**Initial interest debt amount**	0.0023
Youngest date of company creation	0.0018	**GCPI annual**	0.0007	**GDP FTEE quarter**	0.0034	**# mobile phone numbers**	0.0006
Calls ever flag	0.0008	**CCI actual deciles**	0.0003	**CCI expectations**	0.0019	**Calls l90d flag**	0.0006
AR annual	0.0000	**Initial interest debt amount**	0.0000	**Product type grouped**	0.0016	**UCPI annual deciles**	0.0004
GDP FTEE quarter	0.0000	**UCPI annual deciles**	0.0000	**UCPI monthly deciles**	0.0013	**Calls l30d flag**	0.0000

GCPI deciles	0.0000	**Initial total debt amount**	0.0000	**Oldest date of company creation**	0.0000	**Calls ever flag**	0.0000
CCI actual deciles	0.0000	**GDP FTEE annual**	0.0000	**Region**	0.0000	**# phone numbers**	0.0000
GCPI annual deciles	0.0000			**GDP FTEE annual deciles**	0.0000		
CCI actual	0.0000			**UR annual deciles**	0.0000		
				GCPI annual deciles	0.0000		
				Youngest date of company creation	0.0000		
				UCPI annual deciles	0.0000		

Source: Own elaboration

In the reduced models for SMEs, unlike those for Individuals, all categories of determinants remain significant. These include debt characteristics, collection activities, macroeconomic factors, and debtor characteristics. Notably, the Guarantor flag and Judicial flag remain the most influential determinants across most models, with their mixed weights ranging from 0.7012 to 0.8662. Table 66 presents a comprehensive list of the variables used in the reduced models for SMEs, along with the respective weights assigned to each variable within the constructed models.

Table 66. Model variables and their weight for SMEs reduced models

Global		Pre-COVID		COVID		Post-COVID	
Variable	**Weight**	**Variable**	**Weight**	**Variable**	**Weight**	**Variable**	**Weight**
Guarantor flag	0.5130	**Guarantor flag**	0.4256	**Guarantor flag**	0.4605	**Guarantor flag**	0.4763
Judicial flag	0.3532	**Judicial flag**	0.3037	**Judicial flag**	0.3594	**Judicial flag**	0.2249
Product type grouped	0.0457	**Monthly debt amount grouped**	0.0895	**Initial principal debt amount**	0.1068	**RPC ever flag**	0.2000
GDP MP annual	0.0233	**Product type grouped**	0.0505	**Creditor grouped**	0.0187	**UCPI annual**	0.0252
Creditor grouped	0.0184	**# phone numbers**	0.0446	**CCI expectations deciles**	0.0160	**Product type grouped**	0.0236
Monthly debt	0.0178	**CCI expectations deciles**	0.0220	**CCI**	0.0137	**GCPI monthly**	0.0189

amount grouped							
# mobile phone numbers	0.0140	**Calls ever flag**	0.0204	**GDP MP quarter deciles**	0.0082	**Creditor grouped**	0.0166
Initial interest debt amount	0.0112	**Region**	0.0201	**CCI actual deciles**	0.0078	**# fixed line phone numbers**	0.0095
# phone numbers	0.0032	**GCPI monthly**	0.0134	**GDP MP annual deciles**	0.0046	**GCPI annual**	0.0050
		Creditor grouped	0.0088	**Initial interest debt amount**	0.0044	**Monthly debt amount grouped**	0.0000
		CCI actual	0.0014	**UR annual**	0.0000		

Source: Own elaboration

4.3.2.- Summary of Gini and K-S Indicators of the Constructed Models

In this sub-section, we present the Gini and K-S indicators for the models developed for Individuals and SMEs, encompassing both complete and reduced models. This assessment aims to evaluate their predictive power and robustness. The findings are encouraging, as the Gini and K-S indicators align favourably for both Individuals and SMEs when compared to the results found in the literature (as detailed in sub-section 4.2.3 of this document), with Gini indicators ranging from 0.522 to 0.713 in the reduced models and from 0.571 to 0.714 in the complete models, and K-S indicators from 0.379 to 0.577 in the reduced models and from 0.427 to 0.573 in the complete models. It is particularly noteworthy that our research is based on written-off portfolios with average debt ages of 11 years for Individuals and 12 years for SMEs, which typically exhibit lower predictive performance than younger debt portfolios referenced in the literature.

Furthermore, we find that the reduced models for both Individuals and SMEs demonstrate Gini and K-S indicators that are very similar to those of the complete models, albeit with slight reductions. These reduced models are not only simpler and easier to interpret, but they also require less maintenance, as the number of variables has been reduced from the range of 19-26 in the complete models (22-26 for Individuals and 19-24 for SMEs) to just 7-11 variables in the reduced models (7-10 for Individuals and 9-11 for SMEs). Tables 67, 68, 69, and 70 present the Gini and K-S indicators for the complete and reduced models for Individuals and SMEs, respectively.

Table 67. Gini and K-S indicators for Individuals complete models

Models	Gini		K-S	
	Training sample	Testing sample	Training sample	Testing sample
Ind. Global	0.635	0.638	0.485	0.489
Ind. Pre-COVID	0.571	0.562	0.427	0.423
Ind. COVID	0.611	0.611	0.467	0.477
Ind. Post-COVID	0.714	0.714	0.573	0.577

Source: Own elaboration

Table 68. Gini and K-S indicators for Individuals reduced models

Models	Gini		K-S	
	Training sample	Testing sample	Training sample	Testing sample
Ind. Global	0.617	0.620	0.483	0.486
Ind. Pre-COVID	0.532	0.522	0.387	0.379
Ind. COVID	0.599	0.596	0.465	0.468
Ind. Post-COVID	0.711	0.713	0.572	0.577

Source: Own elaboration

Table 69. Gini and K-S indicators for SMEs complete models

Models	Gini		K-S	
	Training sample	Testing sample	Training sample	Testing sample
SME Global	0.662	0.639	0.529	0.515
SME Pre-COVID	0.674	0.610	0.532	0.492
SME COVID	0.677	0.646	0.542	0.514
SME Post-COVID	0.699	0.666	0.550	0.522

Source: Own elaboration

Table 70. Gini and K-S indicators for SMEs reduced models

Models	Gini		K-S	
	Training sample	Testing sample	Training sample	Testing sample
SME Global	0.657	0.633	0.525	0.509
SME Pre-COVID	0.668	0.606	0.532	0.492
SME COVID	0.666	0.648	0.527	0.514
SME Post-COVID	0.692	0.659	0.541	0.517

Source: Own elaboration

4.3.3.- Reduced Models Detail

4.3.3.1.- Individuals Global Model (reduced)

Variables included in the model and their importance as predictor

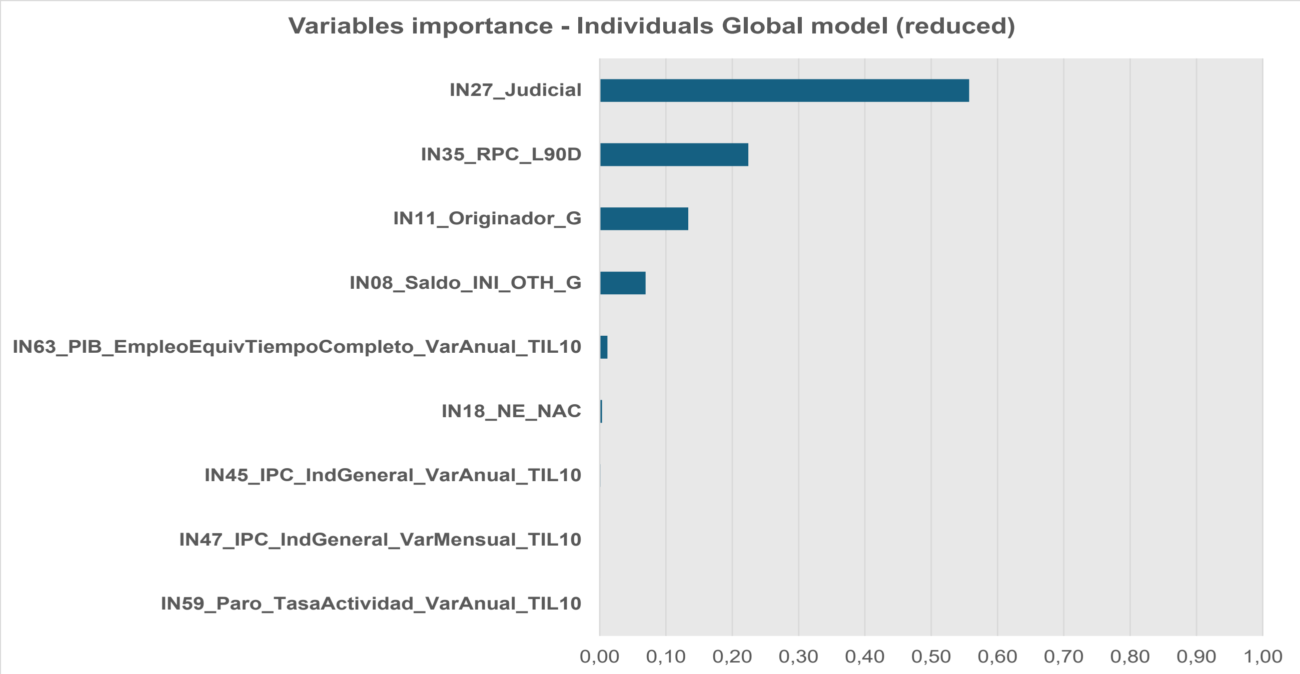
The variables included in this model are the following 9, listed in Table 71 in descending order of importance.

Table 71. Variables included in the Individuals Global model (reduced)

Variable	Description
"IN27_Judicial"	Judicial flag
"IN35_RPC_L90D"	Right-party contact in the last 90 days flag
"IN11_Originador_G"	Creditor grouped
"IN08_Saldo_INI_OTH_G"	Initial other debt amount
"IN63_PIB_EmpleoEquivTiempoCompleto_VarAnual_TIL10"	GDP Full-Time Equivalent Employment on annual variation grouped in deciles
"IN18_NE_NAC"	National nationality flag
"IN45_IPC_IndGeneral_VarAnual_TIL10"	General Consumer Price Index on annual variation grouped in deciles
"IN47_IPC_IndGeneral_VarMensual_TIL10"	General Consumer Price Index on monthly variation grouped in deciles
"IN59_Paro_TasaActividad_VarAnual_TIL10"	Activity Rate on annual variation grouped in deciles

Source: Own elaboration

Figure 68. Variables importance for the Individuals Global model (reduced)

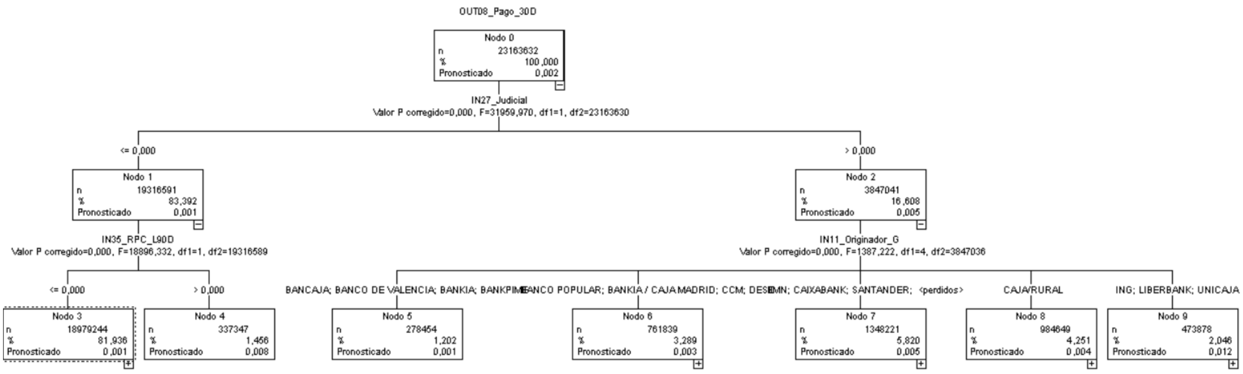


Source: Own elaboration

Model snapshot – initial layers

The following figure shows the initial 3 layers of this model.

Figure 69. Initial layers of the Individuals Global model (reduced)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table, we summarize the Gini and KS indicators for the Individuals Global model.

Table 72. Gini and K-S indicators of the Individuals Global model (reduced)

Indicators	Training sample	Testing sample
Gini	0,617	0,620
K-S	0,483	0,486

Source: Own elaboration

Model predictability: % of accounts and % of payments

For the practical application of each model developed in C&R activities, which is highly focused on resource optimisation, the results are analysed by ranking accounts using the models' scores and the distribution of expected total payments for each score rank. We summarize these analyses in a table and a graphical representation, showing the percentage of accounts included in each score rank and the percentage of payments these accounts represent.

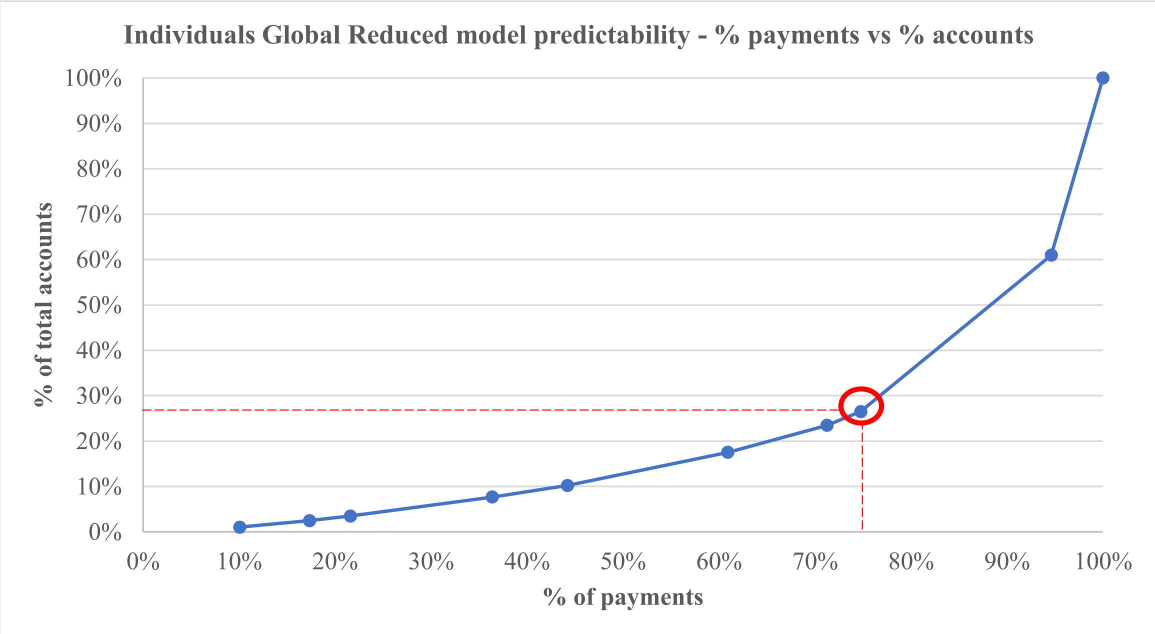
As an illustration, based on the constructed model for Individuals Global (Table 73 and Figure 70), selecting 1% of the accounts identifies 10% of the accounts that will pay in the following 30 days, selecting 10% provides 44% of the payments, and selecting 27% of the accounts yields 75% of the total.

This analysis enables the balancing of management resources, efforts, and costs according to expected future collections' profitability.

A comprehensive analysis has been conducted for each model developed, with a summary of the findings integrated into the respective sections on the models in this document.

Table 73 and Figure 70. Distribution of accounts and payments on the Individuals Global model (reduced)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	95%	61%
0,002	75%	27%
0,003	71%	23%
0,004	61%	18%
0,005	44%	10%
0,006	36%	8%
0,007	22%	4%
0,008	17%	2%
0,017	10%	1%



Source: Own elaboration

4.3.3.2.- Individuals Pre-COVID Model (reduced)

Variables included in the model and their importance as predictor

The variables included in this model are the following 8, as shown in Table 74, ordered in descending importance.

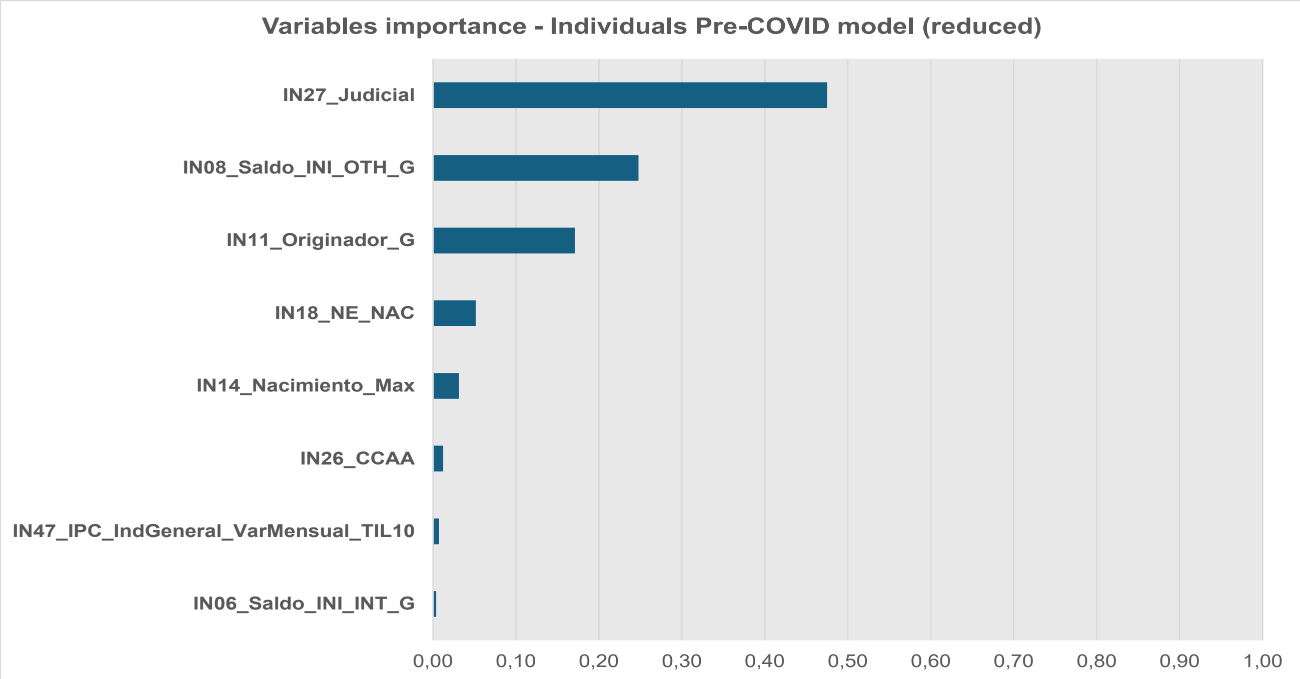
Table 74. Variables included in the Individuals Pre-COVID model (reduced)

Variable	Description
"IN27_Judicial"	Judicial flag
"IN08_Saldo_INI_OTH_G"	Initial other debt amount
"IN11_Originador_G"	Creditor grouped
"IN18_NE_NAC"	National nationality flag
"IN14_Nacimiento_Max"	Youngest date of birth or date of company creation of participants

"IN26_CCAA"	Region
"IN47_IPC_IndGeneral_VarMensual_TIL10"	General Consumer Price Index on monthly variation grouped in deciles
"IN06_Saldo_INI_INT_G"	Initial interest debt amount

Source: Own elaboration

Figure 71. Variables importance for the Individuals Pre-COVID model (reduced)

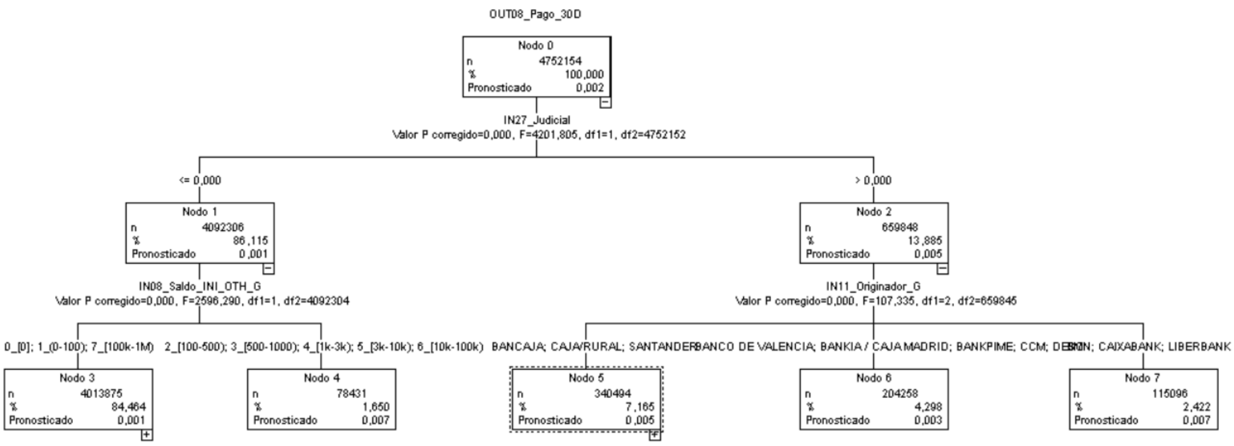


Source: Own elaboration

Model snapshot – initial layers

The following figure shows the initial 3 layers of this model.

Figure 72. Initial layers of the Individuals Pre-COVID model (reduced)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table, we summarize the Gini and KS indicators for the Individuals Pre-COVID model.

Table 75. Gini and K-S indicators of the Individuals Pre-COVID model (reduced)

Indicators	Training sample	Testing sample
Gini	0,532	0,522
K-S	0,387	0,379

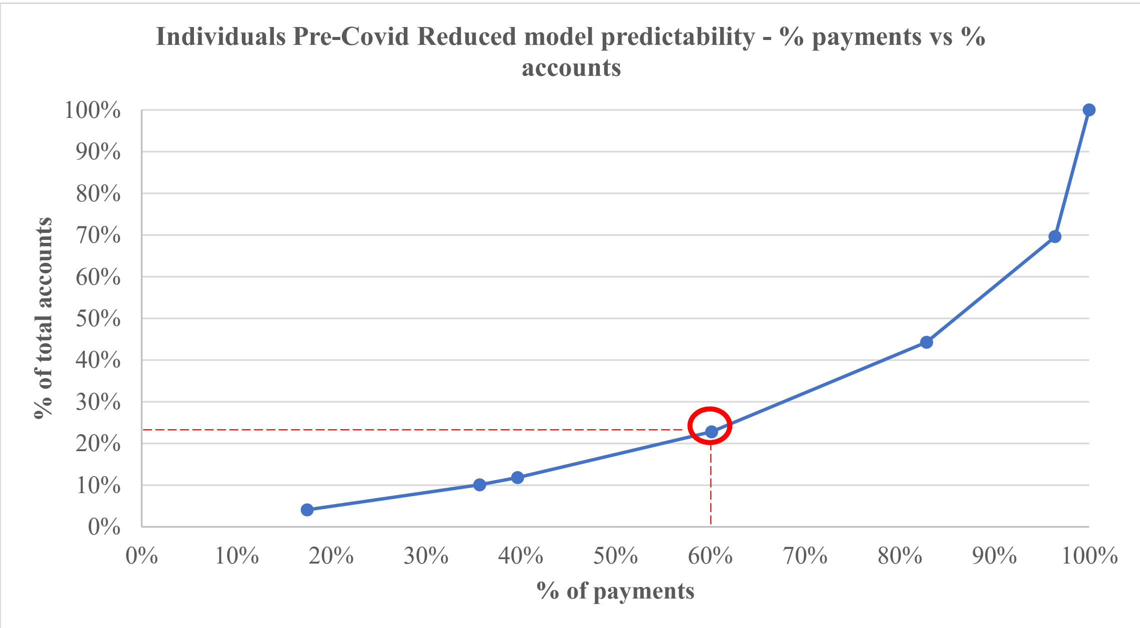
Source: Own elaboration

Model predictability: % of accounts and % of payments

Based on the constructed model for the Individuals Pre-COVID, with a selection of 23% of the accounts, we obtain 60% of the total accounts that will pay in the following 30 days.

Table 76 and Figure 73. Distribution of accounts and payments on the Individuals Pre-COVID model (reduced)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	96%	70%
0,002	83%	44%
0,003	60%	23%
0,004	40%	12%
0,005	36%	10%
0,007	17%	4%



Source: Own elaboration

4.3.3.3.- Individuals COVID Model (reduced)

Variables included in the model and their importance as predictor

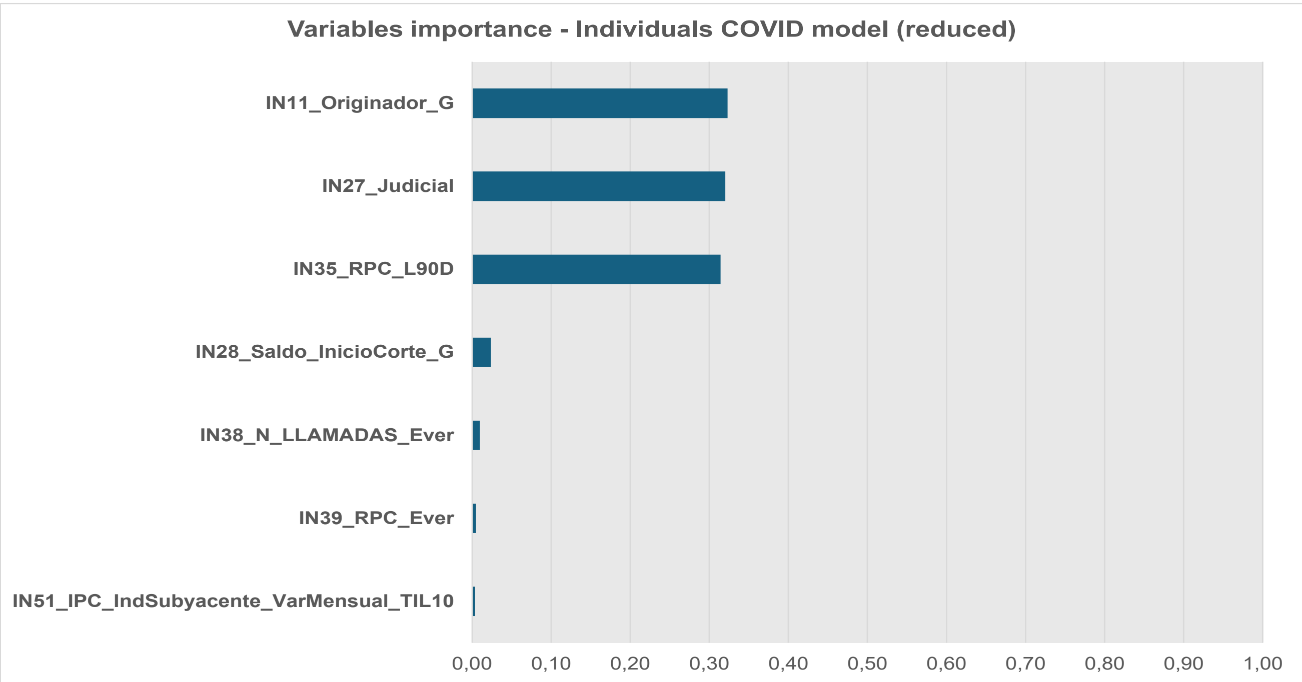
The variables included in this model are the following 7, listed in Table 77 in descending order of importance.

Table 77. Variables included in the Individuals COVID model (reduced)

Variable	Description
"IN11_Originador_G"	Creditor grouped
"IN27_Judicial"	Judicial flag
"IN35_RPC_L90D"	Right-party contact in the last 90 days flag
"IN28_Saldo_InicioCorte_G"	Monthly debt amount grouped
"IN38_N_LLAMADAS_Ever"	Calls ever flag
"IN39_RPC_Ever"	Right-party contact ever flag
"IN51_IPC_IndSubyacente_VarMensual_TIL10"	Underlying Consumer Price Index on monthly variation grouped in deciles

Source: Own elaboration

Figure 74. Variables importance for the Individuals COVID model (reduced)

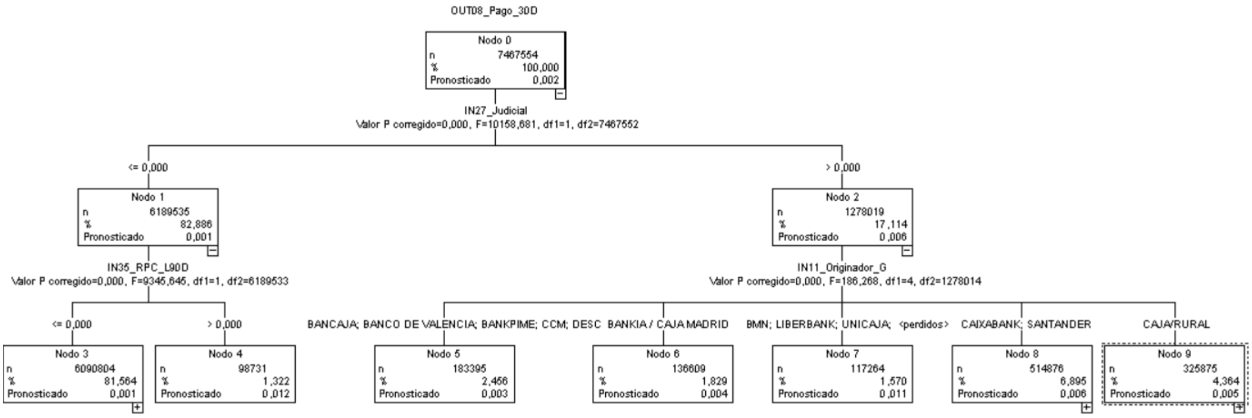


Source: Own elaboration

Model snapshot – initial layers

The following figure shows the initial 3 layers of this model.

Figure 75. Initial layers of the Individuals COVID model (reduced)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table, we summarize the Gini and KS indicators for the Individuals COVID model.

Table 78. Gini and K-S indicators of the Individuals COVID Model (reduced)

Indicators	Training sample	Testing sample
Gini	0,599	0,596
K-S	0,465	0,468

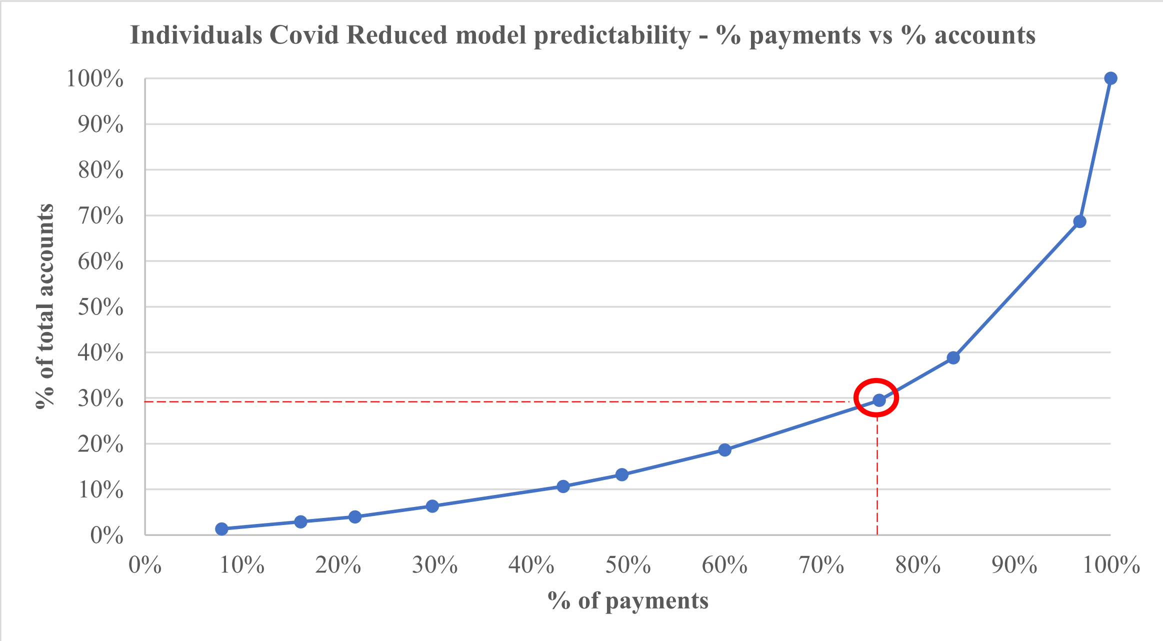
Source: Own elaboration

Model predictability: % of accounts and % of payments

Based on the constructed model for the Individuals COVID, selecting 29% of the accounts yields 76% of the total accounts that will pay in the following 30 days.

Table 79 and Figure 76. Distribution of accounts and payments on the Individuals COVID Model (reduced)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	97%	69%
0,002	84%	39%
0,003	76%	29%
0,004	60%	19%
0,005	49%	13%
0,006	43%	11%
0,007	30%	6%
0,010	22%	4%
0,011	16%	3%
0,012	8%	1%



Source: Own elaboration

4.3.3.4.- Individuals Post-COVID Model (reduced)

Variables included in the model and their importance as predictor

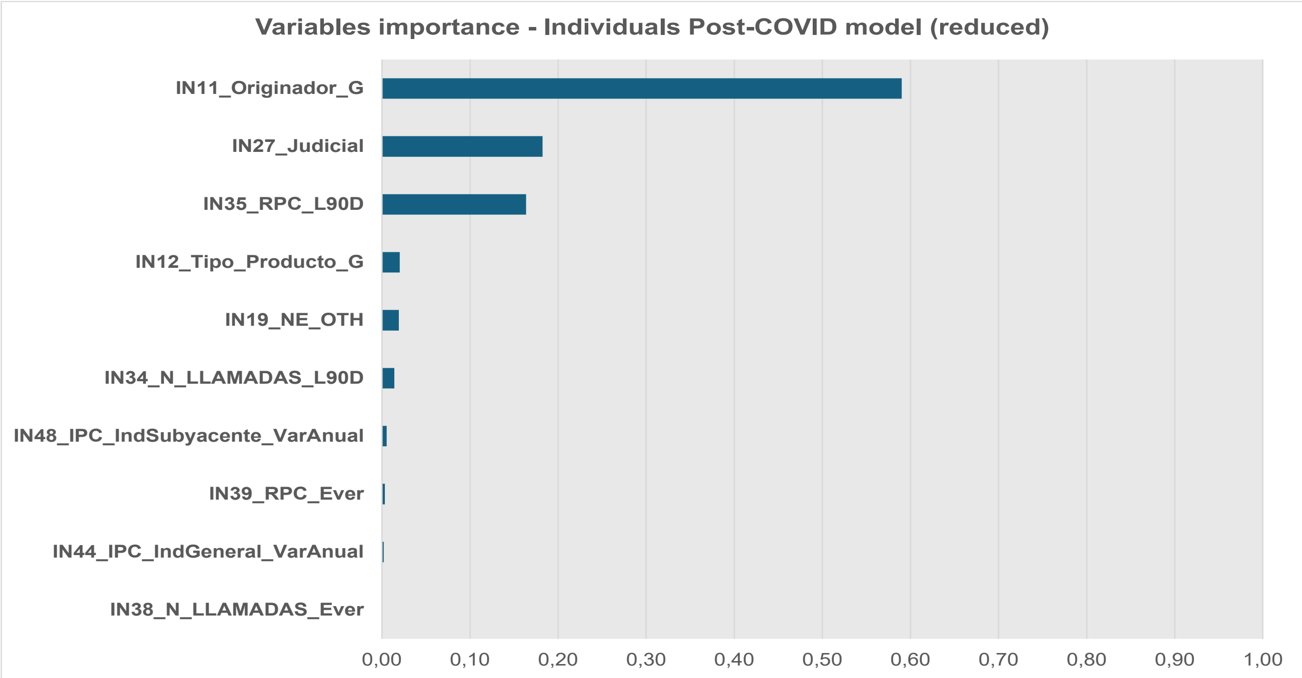
The variables included in this model are the following 10, as shown in Table 80, ordered in descending importance.

Table 80. Variables included in the Individuals Post-COVID model (reduced)

Variable	Description
"IN11_Originador_G"	Creditor grouped
"IN27_Judicial"	Judicial flag
"IN35_RPC_L90D"	Right-party contact in the last 90 days flag
"IN12_Tipo_Producto_G"	Product type grouped
"IN19_NE_OTH"	Other nationality flag
"IN34_N_LLAMADAS_L90D"	Calls in the last 90 days flag
"IN48_IPC_IndSubyacente_VarAnual"	Underlying Consumer Price Index on annual variation
"IN39_RPC_Ever"	Right-party contact ever flag
"IN44_IPC_IndGeneral_VarAnual"	General Consumer Price Index on annual variation
"IN38_N_LLAMADAS_Ever"	Calls ever flag

Source: Own elaboration

Figure 77. Variables importance for the Individuals Post-COVID model (reduced)

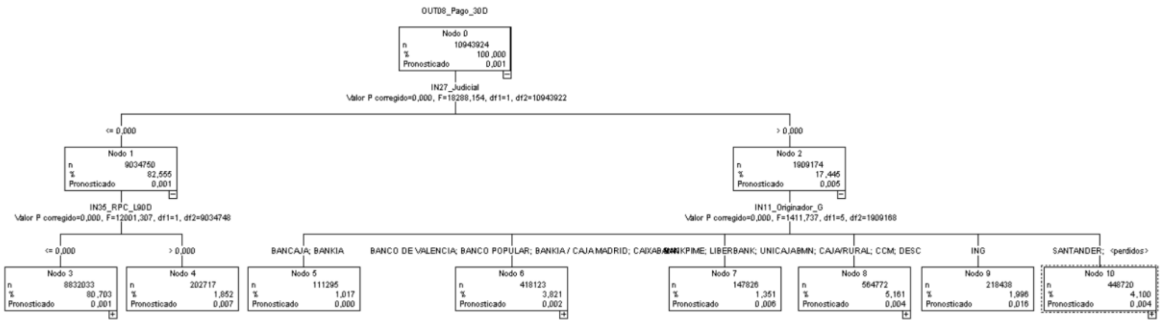


Source: Own elaboration

Model snapshot – initial layers

The following figure shows the initial 3 layers of this model.

Figure 78. Initial layers of the Individuals Post-COVID model (reduced)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table, we summarize the Gini and KS indicators for the Individuals Post-COVID model.

Table 81. Gini and K-S indicators of the Individuals Post-COVID model (reduced)

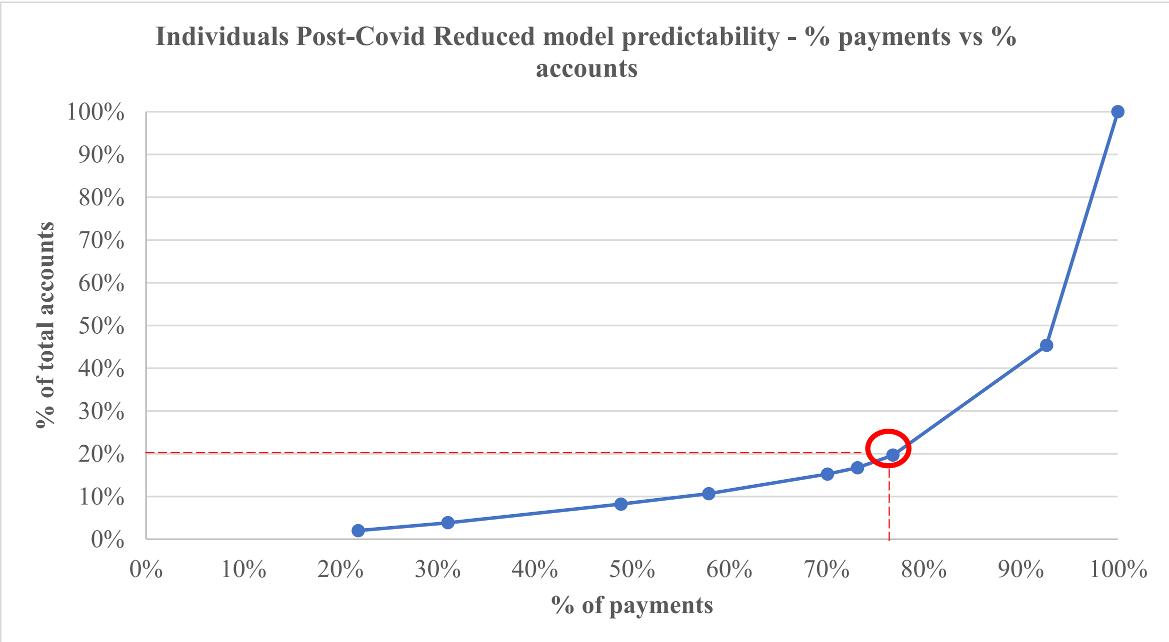
Indicators	Training sample	Testing sample
Gini	0,711	0,713
K-S	0,572	0,577

Source: Own elaboration

Model predictability: % of accounts and % of payments

Based on the constructed model for the Individuals Post-COVID, with a 20% sample, we obtain 77% of the total accounts that will pay in the following 30 days.

Figure 79 and Table 82. Distribution of accounts and payments on the Individuals Post-COVID model (reduced)



MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	93%	45%
0,002	77%	20%
0,003	73%	17%
0,004	70%	15%
0,005	58%	11%
0,006	49%	8%
0,007	31%	4%
0,016	22%	2%

Source: Own elaboration

4.3.3.5.- SMEs Global Model (reduced)

Variables included in the model and their importance as predictor

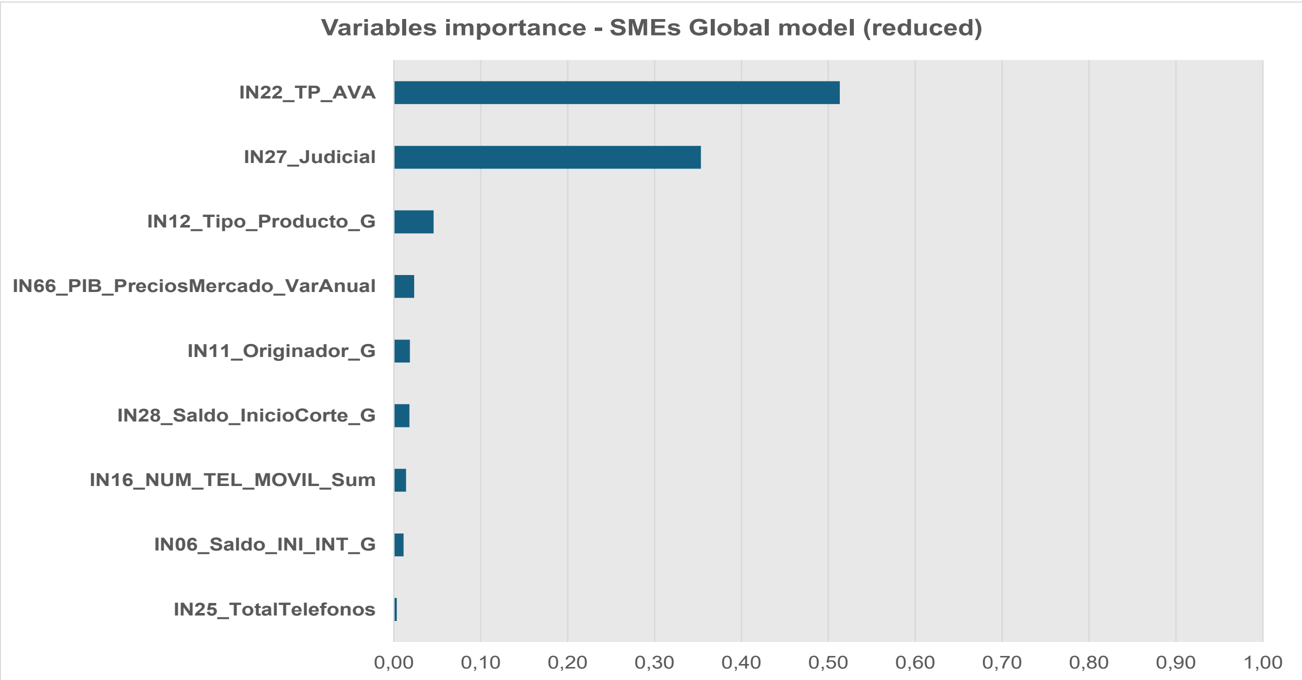
The variables included in this model are the following 9, listed in Table 83 in descending order of importance.

Table 83. Variables included in the SMEs Global model (reduced)

Variable	Description
"IN22_TP_AVA"	Guarantor flag
"IN27_Judicial"	Judicial flag
"IN12_Tipo_Producto_G"	Product type grouped
"IN66_PIB_PreciosMercado_VarAnual"	GDP Market Prices on annual variation
"IN11_Originador_G"	Creditor grouped
"IN28_Saldo_InicioCorte_G"	Monthly debt amount grouped
"IN16_NUM_TEL_MOVIL_Sum"	Number of mobile phone numbers
"IN06_Saldo_INI_INT_G"	Initial interest debt amount
"IN25_TotalTelefonos"	Number of phone numbers

Source: Own elaboration

Figure 80. Variables importance for the SMEs Global model (reduced)

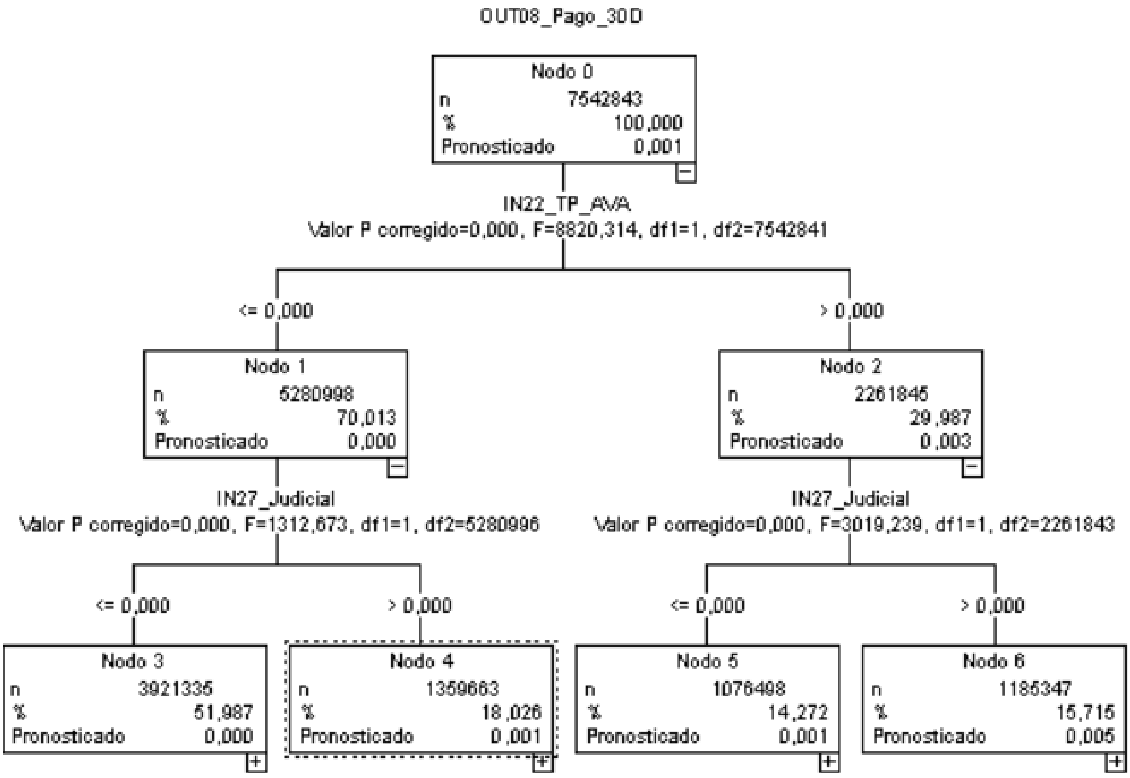


Source: Own elaboration

Model snapshot – initial layers

The following figure shows the initial 3 layers of this model.

Figure 81. Initial layers of the SMEs Global model (reduced)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table, we summarize the Gini and KS indicators for the SMEs Global model.

Table 84. Gini and K-S indicators of the SMEs Global model (reduced)

Indicators	Training sample	Testing sample
Gini	0,657	0,633
K-S	0,525	0,509

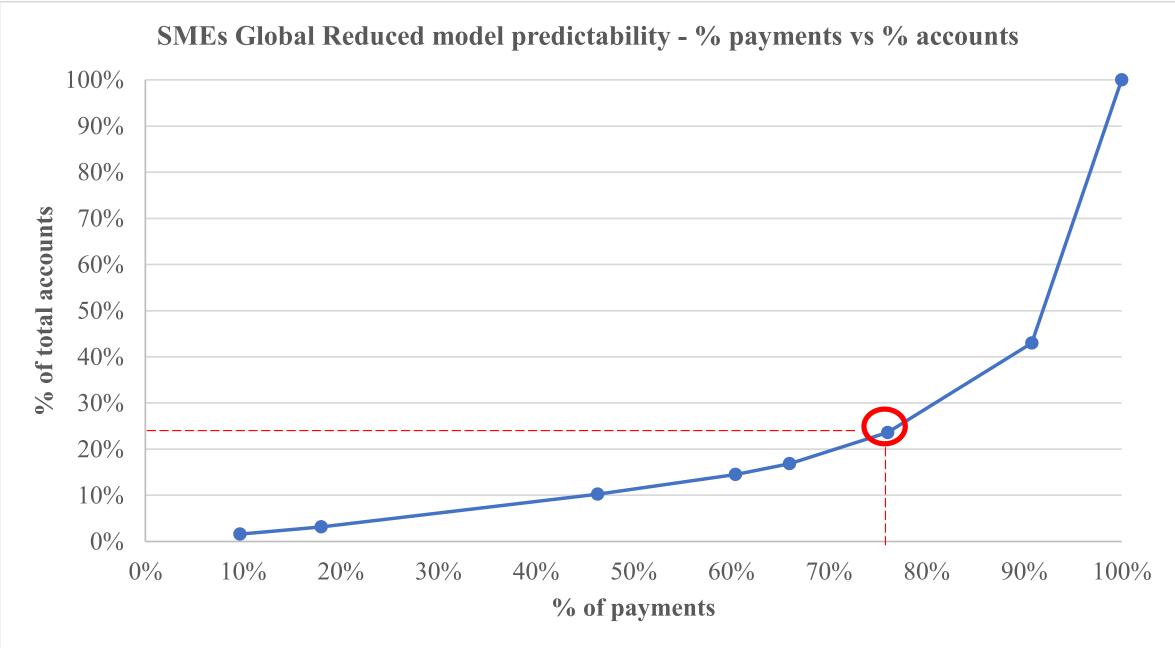
Source: Own elaboration

Model predictability: % of accounts and % of payments

Based on the constructed model for SMEs Global, with a selection of 24% of the accounts, we obtain 76% of the total accounts that will pay in the following 30 days.

Table 85 and Figure 82. Distribution of accounts and payments on the SMEs Global model (reduced)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	91%	43%
0,002	76%	24%
0,003	66%	17%
0,004	60%	15%
0,005	46%	10%
0,007	18%	3%
0,008	10%	2%



Source: Own elaboration

4.3.3.6.- SMEs Pre-COVID Model (reduced)

Variables included in the model and their importance as predictor

The variables included in this model are the following 11, as shown in Table 86, ordered in descending importance.

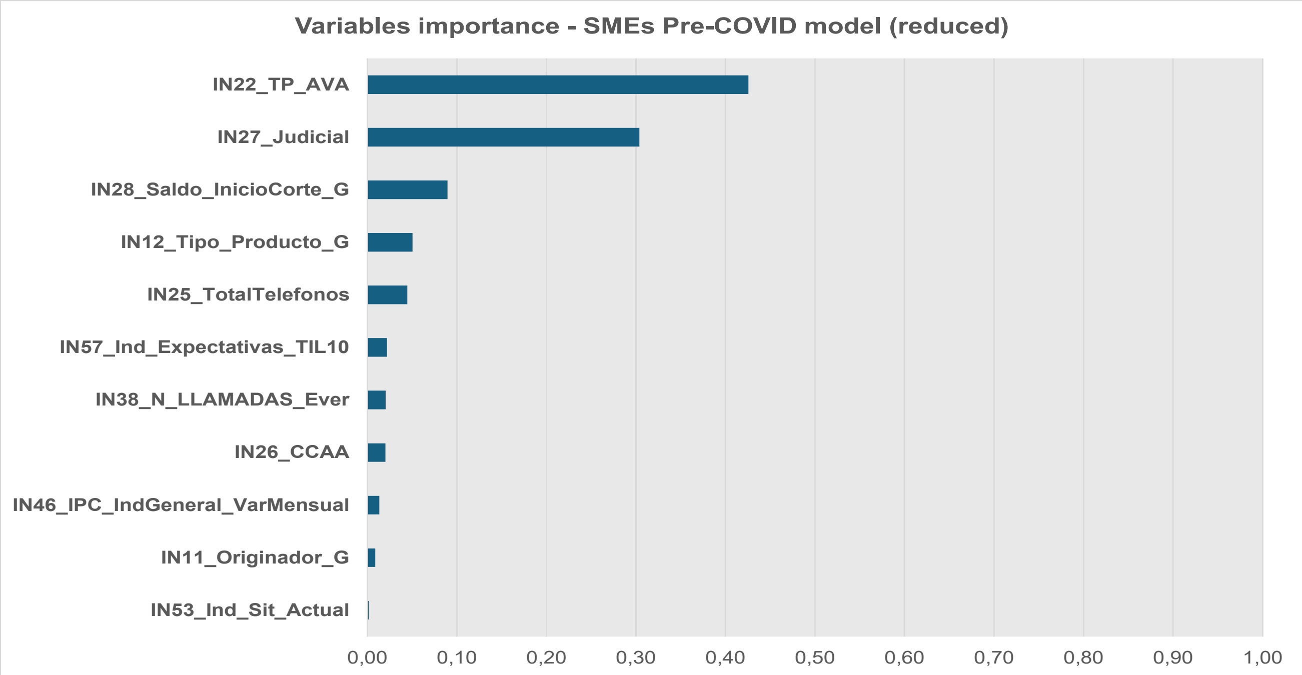
Table 86. Variables included in the SMEs Pre-COVID model (reduced)

Variable	Description
"IN22_TP_AVA"	Guarantor flag
"IN27_Judicial"	Judicial flag
"IN28_Saldo_InicioCorte_G"	Monthly debt amount grouped
"IN12_Tipo_Producto_G"	Product type grouped
"IN25_TotalTelefonos"	Number of phone numbers
"IN57_Ind_Expectativas_TIL10"	Consumer Confidence Index expectations grouped in deciles

"IN38_N_LLAMADAS_Ever"	Calls ever flag
"IN26_CCAA"	Region
"IN46_IPC_IndGeneral_VarMensual"	General Consumer Price Index on monthly variation
"IN11_Originador_G"	Creditor grouped
"IN53_Ind_Sit_Actual"	Consumer Confidence Index actual situation

Source: Own elaboration

Figure 83. Variables importance for the SMEs Pre-COVID model (reduced)

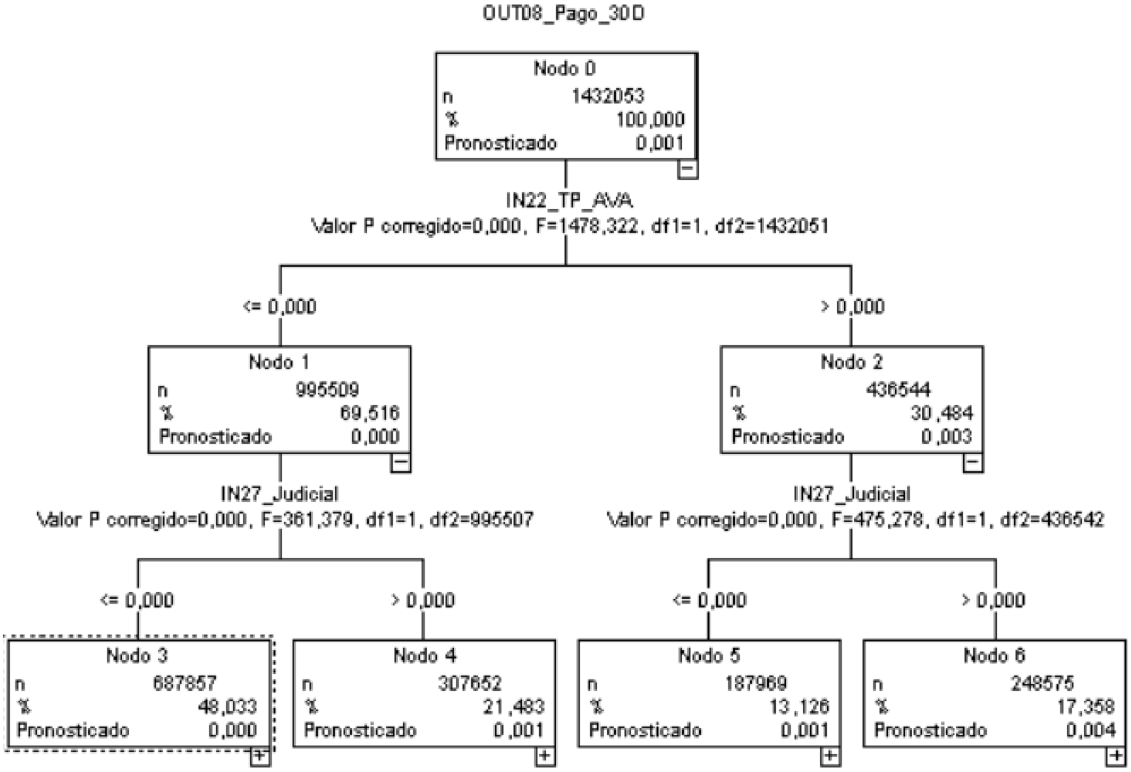


Source: Own elaboration

Model snapshot – initial layers

In the following Figure, the initial 3 layers of this model are shown.

Figure 84. Initial layers of the SMEs Pre-COVID model (reduced)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table, we summarize the Gini and KS indicators for the SMEs Pre-COVID model.

Table 87. Gini and K-S indicators of the SMEs Pre-COVID model (reduced)

Indicators	Training sample	Testing sample
Gini	0,668	0,606
K-S	0,532	0,492

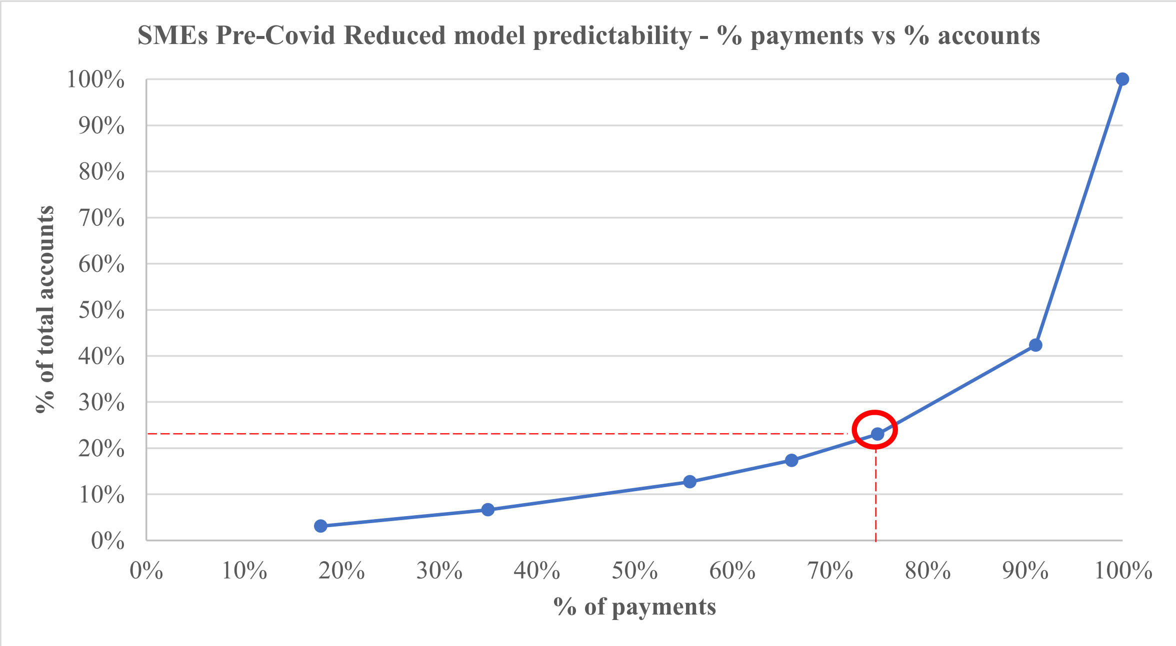
Source: Own elaboration

Model predictability: % of accounts and % of payments

Based on the constructed model for the SMEs Pre-COVID, with a selection of 23% of the accounts, we obtain 75% of the total accounts that will pay in the following 30 days.

Table 88 and Figure 85. Distribution of accounts and payments on the SMEs Pre-COVID model (reduced)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	91%	42%
0,002	75%	23%
0,003	66%	17%
0,004	56%	13%
0,006	35%	7%
0,007	18%	3%



Source: Own elaboration

4.3.3.7.- SMEs COVID Model (reduced)

Variables included in the model and their importance as predictor

The variables included in this model are the following 11, as shown in Table 89, ordered in descending importance.

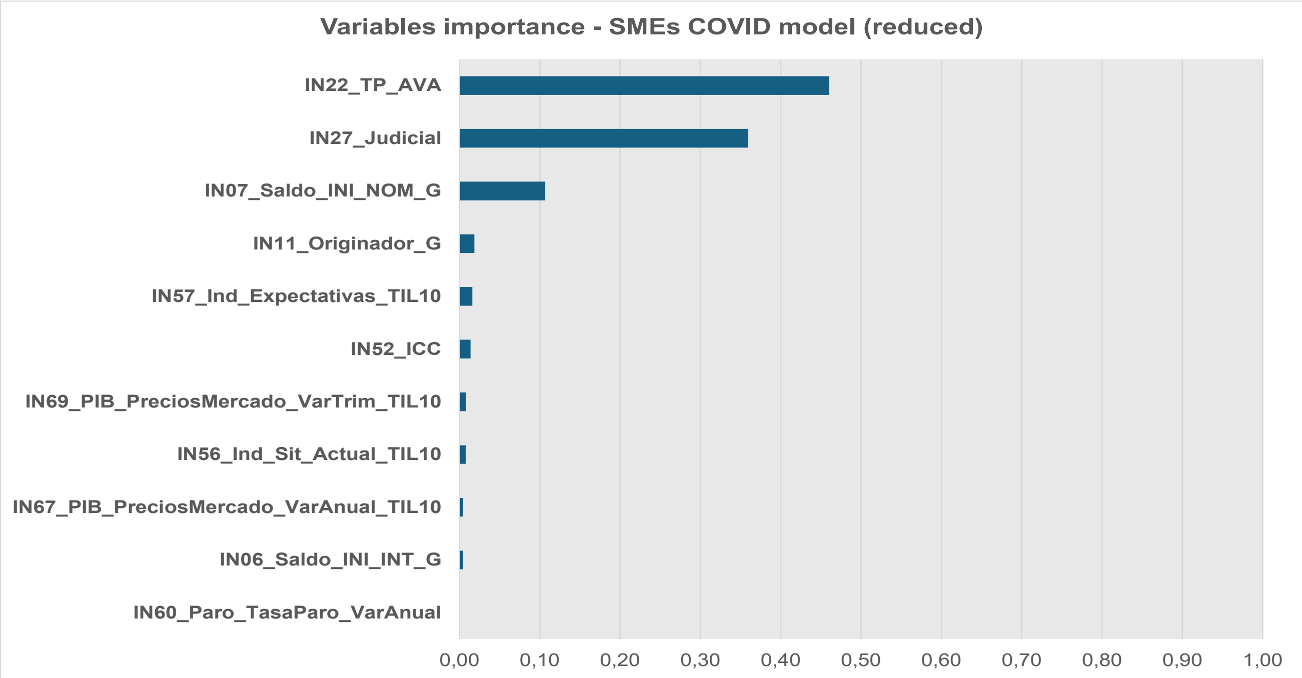
Table 89. Variables included in the SMEs COVID model (reduced)

Variable	Description
"IN22_TP_AVA"	Guarantor flag
"IN27_Judicial"	Judicial flag
"IN07_Saldo_INI_NOM_G"	Initial principal debt amount
"IN11_Originador_G"	Creditor grouped
"IN57_Ind_Expectativas_TIL10"	Consumer Confidence Index expectations grouped in deciles
"IN52_ICC"	Consumer Confidence Index

"IN69_PIB_PreciosMercado_VarTrim_TIL10"	GDP Market Prices on quarterly variation grouped in deciles
"IN56_Ind_Sit_Actual_TIL10"	Consumer Confidence Index actual situation grouped in deciles
"IN67_PIB_PreciosMercado_VarAnual_TIL10"	GDP Market Prices on annual variation grouped in deciles
"IN06_Saldo_INI_INT_G"	Initial interest debt amount
"IN60_Paro_TasaParo_VarAnual"	Unemployment Rate on annual variation

Source: Own elaboration

Figure 86. Variables importance for the SMEs COVID model (reduced)

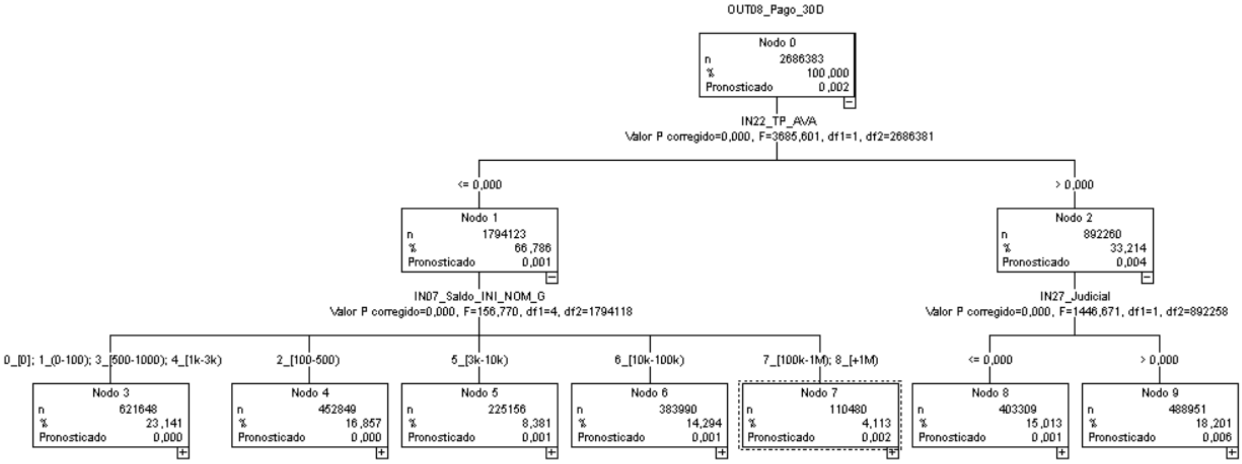


Source: Own elaboration

Model snapshot – initial layers

The following figure shows the initial 3 layers of this model.

Figure 87. Initial layers of the SMEs COVID model (reduced)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table, we summarize the Gini and KS indicators for the SMEs COVID model.

Table 90. Gini and K-S indicators of the SMEs COVID model (reduced)

Indicators	Training sample	Testing sample
Gini	0,666	0,648
K-S	0,527	0,514

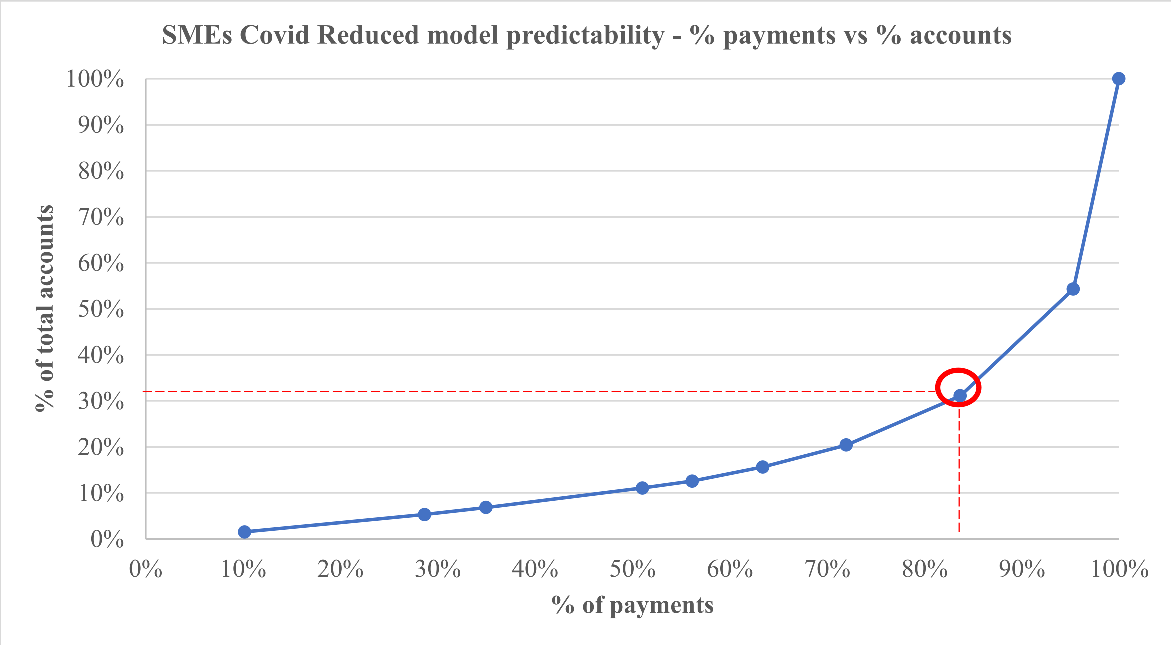
Source: Own elaboration

Model predictability: % of accounts and % of payments

Based on the constructed model for the SME COVID, with a sample of 31% of the accounts, we obtain 84% of the total accounts that will pay in the next 30 days.

Table 91 and Figure 88. Distribution of accounts and payments on the SMEs COVID model (reduced)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	95%	54%
0,002	84%	31%
0,003	72%	20%
0,004	63%	16%
0,005	56%	13%
0,006	51%	11%
0,007	35%	7%
0,008	29%	5%
0,011	10%	2%



Source: Own elaboration

4.3.3.8.- SMEs Post-COVID model (reduced)

Variables included in the model and their importance as predictor

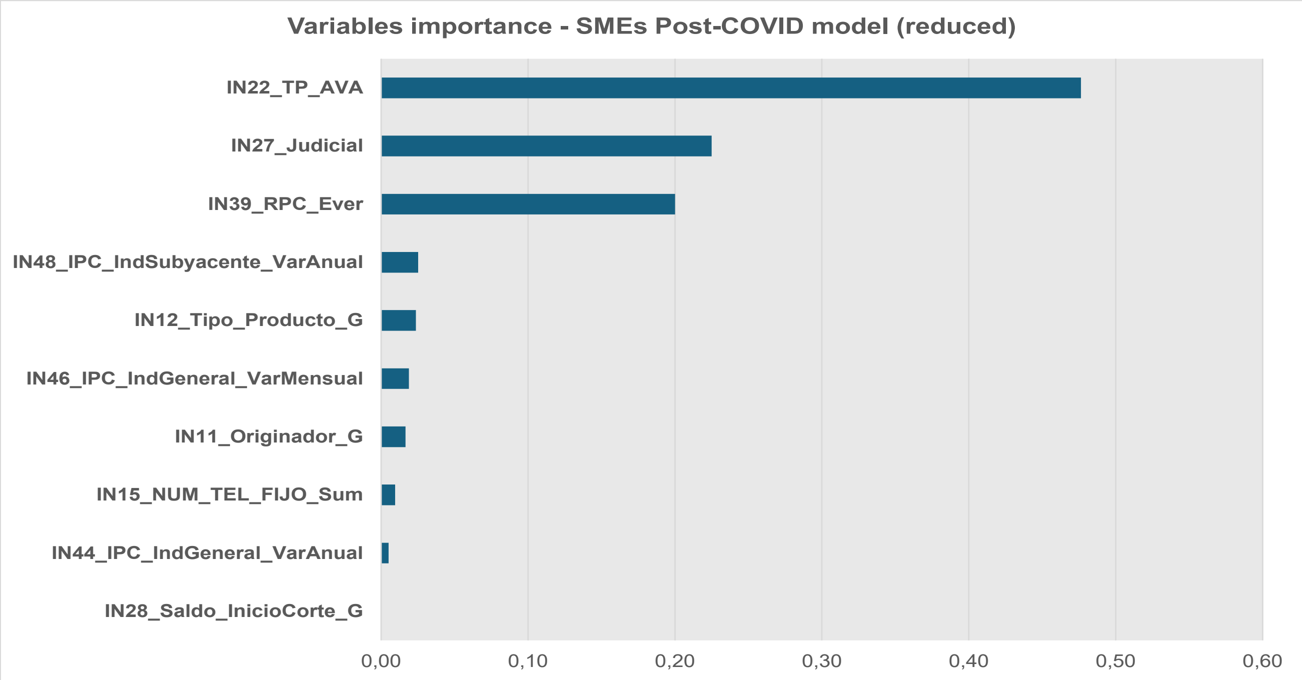
The variables included in this model are the following 10 included in Table 92 ordered in descendent importance.

Table 92. Variables included in the SMEs Post-COVID model (reduced)

Variable	Description
"IN22_TP_AVA"	Guarantor flag
"IN27_Judicial"	Judicial flag
"IN39_RPC_Ever"	Right-party contact ever flag
"IN48_IPC_IndSubyacente_VarAnual"	Underlying Consumer Price Index on annual variation
"IN12_Tipo_Producto_G"	Product type grouped
"IN46_IPC_IndGeneral_VarMensual"	General Consumer Price Index on monthly variation
"IN11_Originador_G"	Creditor grouped
"IN15_NUM_TEL_FIJO_Sum"	Number of fixed line phone numbers
"IN44_IPC_IndGeneral_VarAnual"	General Consumer Price Index on annual variation
"IN28_Saldo_InicioCorte_G"	Monthly debt amount grouped

Source: Own elaboration

Figure 89. Variables importance for the SMEs Post-COVID model (reduced)

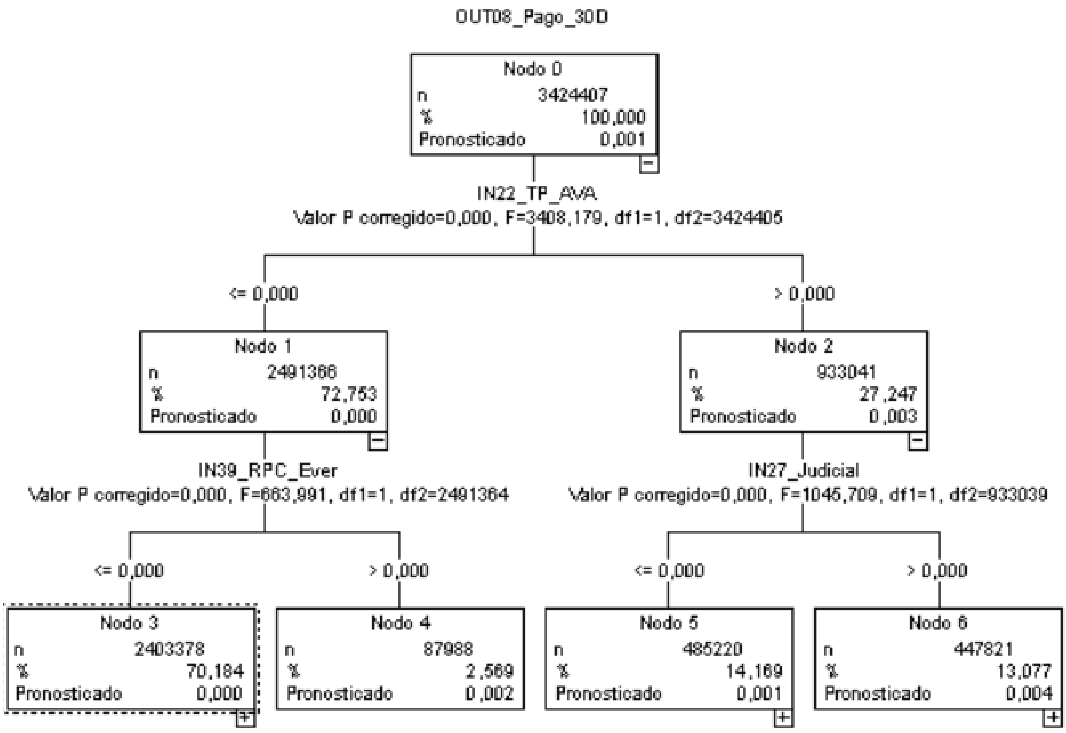


Source: Own elaboration

Model snapshot – initial layers

The following figure shows the initial 3 layers of this model.

Figure 90. Initial layers of the SMEs Post-COVID model (reduced)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table, we summarize the Gini and KS indicators for the SMEs post-COVID model.

Table 93. Gini and K-S indicators of the SMEs Post-COVID model (reduced)

Indicators	Training sample	Testing sample
Gini	0,692	0,659
K-S	0,541	0,517

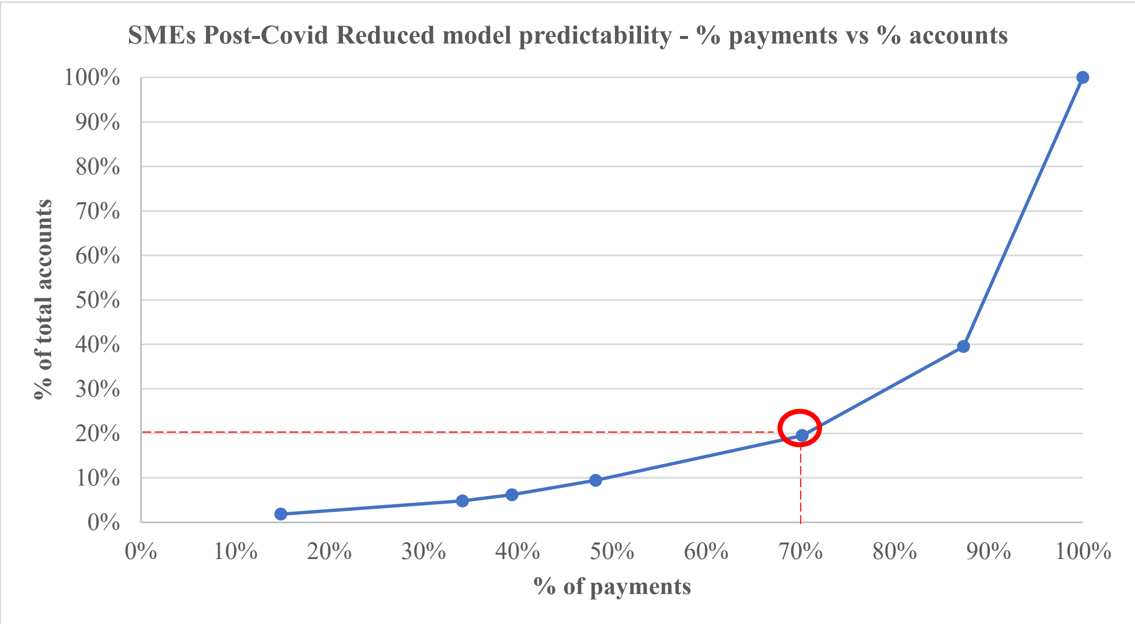
Source: Own elaboration

Model predictability: % of accounts and % of payments

Based on the constructed model for the SMEs post-COVID, with a selection of 20% of the accounts, we obtain 70% of the total accounts that will pay in the following 30 days.

Table 94 and Figure 91. Distribution of accounts and payments on the SMEs Post-COVID model (reduced)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	87%	39%
0,002	70%	20%
0,003	48%	9%
0,004	39%	6%
0,006	34%	5%
0,009	15%	2%



Source: Own elaboration

SECTION 5.- CONCLUSIONS AND RECOMMENDATIONS

5.1.- RESULTS SUMMARY AND CONFIRMATION OF THE RESEARCH QUESTION

We developed statistical models to predict recovery rates for unsecured debt written-off portfolios acquired from banks and financial institutions, distinguishing between Individual and SME cases over a global observation period from 2018 to 2023. Our analysis covers four distinct timeframes: the Global period (2018-2023), the Pre-COVID period (2018-2019), the COVID period (2020-2021), and the Post-COVID period (2022-2023), following an approach to the timing split similar to the one found in the literature (Xu et al., 2024). The models demonstrated strong predictive capabilities, as evidenced by the comparison of their Gini and K-S indicators with results found in the existing literature; with Gini coefficients in the range of 0.522-0.713 and K-S indicators in the range of 0.379-0.577 in our models, and best results found in the literature of Gini coefficient of 0.7076 and K-S indicator of 0.5647 (as detailed in sub-sections 4.2.3 and 4.3.2 of this document; Dumitrescu et al., 2022).

These models incorporated statistically relevant variables from four primary categories: debt characteristics, debtor characteristics, collection activities, and macroeconomic factors, consistent with findings in the related literature. Our modelling approach is consistent with the academic literature evolution and recent developments related to the mentioned four data sources (Xu et al., 2024; Li et al., 2024; Grunert & Weber, 2005).

Initially, the models included all variables that were statistically significant. However, it became clear that many of the variables exerted only a marginal influence on the models for both Individuals and SMEs. As a result, we developed a streamlined version of the models that focuses solely on the more impactful variables. These simplified models, which utilize a reduced number of variables (7-11 compared to the 19-26 in the complete models), maintain comparable predictive capabilities in terms of Gini and K-S indicators while offering easier interpretation and maintenance.

In the following tables, we present a summary of the variables included in the models we constructed, along with their statistical significance. Additionally, a "heat" map has been created to visually illustrate the weight of these variables and their evolution over the various periods. Furthermore, we have categorized the weight of the variables according to the four types considered: C&R activities, debt characteristics, debtor characteristics, and macroeconomic factors.

Table 95. Summary of variables weight and "heat" map of the Individuals models

Pre-COVID		COVID		Post-COVID	
Variable	Weight	Variable	Weight	Variable	Weight
Judicial flag	0,4752	Creditor grouped	0,3230	Creditor grouped	0,5901
Initial other debt amount	0,2476	Judicial flag	0,3202	Judicial flag	0,1824
Creditor grouped	0,1709	RPC 190d	0,3143	RPC 190d	0,1636
National nationality flag	0,0514	Monthly debt amount grouped	0,0239	Product type grouped	0,0201
Youngest DOB	0,0314	Calls ever flag	0,0097	Other nationality flag	0,0192
Region	0,0122	RPC ever flag	0,0050	Calls 190d flag	0,0140
GCPI monthly deciles	0,0075	UCPI monthly deciles	0,0039	UCPI annual	0,0054
Initial interest debt amount	0,0039			RPC ever flag	0,0032
				GCPI annual	0,0020
				Calls ever flag	0,0000

Source: Own elaboration

Table 96. Summary of variables weight and "heat" map of the SMEs models

Pre-COVID		COVID		Post-COVID	
Variable	Weight	Variable	Weight	Variable	Weight
Guarantor flag	0,4256	Guarantor flag	0,4605	Guarantor flag	0,4763
Judicial flag	0,3037	Judicial flag	0,3594	Judicial flag	0,2249
Monthly debt amount grouped	0,0895	Initial principal debt amount	0,1068	RPC ever flag	0,2000
Product type grouped	0,0505	Creditor grouped	0,0187	UCPI annual	0,0252
# phone numbers	0,0446	CCI expectations deciles	0,0160	Product type grouped	0,0236
CCI expectations deciles	0,0220	CCI	0,0137	GCPI monthly	0,0189
Calls ever flag	0,0204	GDP MP quarter deciles	0,0082	Creditor grouped	0,0166
Region	0,0201	CCI actual deciles	0,0078	# fixed line phone numbers	0,0095
GCPI monthly	0,0134	GDP MP annual deciles	0,0046	GCPI annual	0,0050
Creditor grouped	0,0088	Initial interest debt amount	0,0044	Monthly debt amount grouped	0,0000
CCI actual	0,0014	UR annual	0,0000		

Source: Own elaboration

Table 97. Summary of variables weight by variable type of the Individuals models

Pre-COVID		COVID		Post-COVID	
Type of variable	Weight	Type of variable	Weight	Type of variable	Weight
C&R activities	0,4752	C&R activities	0,6492	C&R activities	0,3632
Debt characteristics	0,4224	Debt characteristics	0,3469	Debt characteristics	0,6102
Debtor characteristics	0,0950	Debtor characteristics	0,0000	Debtor characteristics	0,0192
Macroeconomic	0,0075	Macroeconomic	0,0039	Macroeconomic	0,0074

Source: Own elaboration

Table 98. Summary of variables weight by variable type of the SMEs models

Pre-COVID		COVID		Post-COVID	
Type of variable	Weight	Type of variable	Weight	Type of variable	Weight
C&R activities	0,3687	C&R activities	0,3594	C&R activities	0,4344
Debt characteristics	0,5744	Debt characteristics	0,5904	Debt characteristics	0,5165
Debtor characteristics	0,0201	Debtor characteristics	0,0000	Debtor characteristics	0,0000
Macroeconomic	0,0368	Macroeconomic	0,0503	Macroeconomic	0,0491

Source: Own elaboration

The aim of this research is to identify the key determinants of RR for unsecured debt portfolios that have been written off and subsequently acquired from banks and financial institutions. Additionally, it seeks to evaluate the impact of the COVID-19 pandemic (and the measures delivered to mitigate its effects) on these determinants: which variables significantly predict short-term repayment behaviour before, during, and after the COVID-19 crisis? By analysing the available data and developed models for both Individuals and SMEs, we identify the discriminant variables affecting recovery rates. Our findings confirm that these variables have been influenced by the COVID-19 pandemic (and its linked measures) in both the Individual and SME segments. The pandemic clearly altered the significance of these variables, with a more pronounced effect observed in the Individuals models.

5.2.- GENERAL RESULTS OF THE RESEARCH

In the analysis of Individuals models, notable distinctions are observed across different time frames. In the Pre-COVID model, the Judicial flag emerges as the most significant variable, with a weight of 0.4752, followed by the Initial other debt amount at 0.2476, and the Creditor variable at 0.1709. Collectively, these three variables account for a substantial cumulative weight of 0.8937. Subsequent variables that pertain to debtor characteristics—including the National nationality flag (0.0514), the Youngest date of birth of a participant (0.0314), and Region (0.0122)—are also integrated into the model. Furthermore, the variables GCPI and Initial interest debt amount are included; however, their contributions are marginal, with each exhibiting a weight lower than 0.0100.

In contrast, the COVID Individuals models illustrate a shift in variable importance. Here, three variables gain prominence with closely comparable weights: Creditor (0.3230), Judicial flag (0.3202), and RPC last 90 days (0.3143), together constituting a cumulative weight of 0.9575. Following these, the Monthly debt amount emerges as the next significant variable with a weight of 0.0239. Additional variables such as Calls ever flag, RPC ever flag, and UCPI are also incorporated, albeit with marginal importance. The COVID model reveals an increased emphasis on variables related to collections activities—particularly the RPC last 90 days—alongside a persistent relevance of the Judicial flag. Notably, there is a reduction in the representation of variables associated with debtor characteristics compared to the Pre-COVID model.

Turning to the Post-COVID Individuals models, the three primary variables maintain their significance, though there is a notable shift in their respective weights: Creditor (0.5901), Judicial flag (0.1824), and RPC last 90 days (0.1636) result in a cumulative weight of 0.9361 for these key determinants. Subsequent variables include Product type (0.0201), Other nationality flag (0.0192), and Calls last 90 days (0.0140). Variables such as UCPI, RPC ever flag, GCPI, and Calls ever flag continue to feature within the model but demonstrate marginal significance.

This comparative analysis underscores the evolving landscape of relevant variables within Individuals models across different time periods, highlighting shifts in focus and weight attributed to factors linked with debt and debtor characteristics, and collection activities.

Within the constructed models for SMEs, a consistent observation is made regarding the dominance of two primary variables throughout the defined cycle: the Guarantor flag and the Judicial flag. Notably, while these variables retain their central role, their respective weights exhibit variability across different models. Moreover, the four categories of variables utilized in the modelling process—namely, debt characteristics, debtor characteristics, collection activities, and macroeconomic factors—demonstrate varying degrees of significance.

In the Pre-COVID SMEs model, the most prominent variables are identified as the Guarantor flag (weight of 0.4256), the Judicial flag (weight of 0.3037), and the Monthly debt amount (weight of 0.0895). Collectively, these three variables contribute a cumulative weight of 0.8188. Following these, a range of additional variables associated with debt characteristics, collection activities, macroeconomic factors, and debtor characteristics are incorporated into the model. Specifically, these include the Product type (weight of 0.0505), Number of phone numbers (weight of 0.0446), CCI expectations (weight of 0.0220), Calls ever flag (weight of 0.0204), Region (weight of 0.0201), and GCPI (weight of 0.0134). Furthermore, the variables Creditor and CCI actual display marginal significance within this model, each exhibiting a weight lower than 0.0100. This model is notable for possessing a relatively high number of variables with substantial weights (exceeding 0.0100).

Transitioning to the SMEs COVID model, a notable shift in variable prominence is observed. The Guarantor flag (weight of 0.4605), the Judicial flag (weight of 0.3594), and the Initial principal debt amount (weight of 0.1068) emerge as the three most relevant variables, together accounting for a cumulative weight of 0.9267. This marks a significant increase in the weights assigned to these core variables when compared to the Pre-COVID model (0.9267 vs. 0.8188). Subsequent variables in this model include those pertaining to debt characteristics and macroeconomic factors, such as Creditor (weight of 0.0187), CCI expectations (weight of 0.0160), and CCI (weight of 0.0137). While additional variables are incorporated, they possess marginal weights in the model, including GDP MP quarter and annual, CCI actual, Initial interest debt amount, and UR annual. Importantly, a discernible decline in the weight of the collection activities variables is evident when compared to the Pre-COVID model.

Lastly, the SMEs Post-COVID model highlights the Guarantor flag (weight of 0.4763), the Judicial flag (weight of 0.2249), and the RPC ever flag (weight of 0.2000) as the primary variables. Together, these three account for a cumulative weight of 0.9012. Relevant additional variables in this model include UCPI annual (weight of 0.0252), Product type (weight of 0.0236), GCPI monthly (weight of 0.0189), and Creditor (weight of 0.0166). To complement the model, variables such as the Number of fixed line numbers, GCPI annual, and Monthly debt amount are included, albeit with marginal weight in the overall analysis.

5.3.- DISCUSSION OF THE RESEARCH RESULTS

The four primary sources of information identified in the literature for LGD and RR modelling, which were utilized in our research, include borrower characteristics, debt characteristics, macroeconomic variables, and C&R activities. All four sources have been confirmed as relevant for our modelling efforts in both the Individuals and SMEs segments, as found in recent developments of academic literature (Xu et al., 2024; Li et al., 2024; Li et al., 2023A; Nazemi at al., 2022); though there are notable disparities between the two segments and some differences when compared to previous academic studies.

In the literature, borrower characteristics and debt characteristics are the most commonly discussed sources (Calabrese & Zanin, 2022; Chalupka & Kopecsni, 2008), with macroeconomic variables occasionally included (Nazemi et al., 2022; Wang et al., 2020). More recently, some authors have begun to consider repayment behaviour and C&R variables as well (Xu et al., 2024; Li et al., 2024; Li et al., 2023A). Our models place the greatest emphasis on debt characteristics and C&R variables, for both Individuals and SMEs. This approach contrasts with traditional literature, which typically prioritizes borrower characteristics and debt characteristics.

It is particularly noteworthy that borrower characteristics have a limited presence in our models: for the SMEs model, only the Region was included in the Pre-COVID model, with no other borrower variables featured. In the Individuals model, Nationality, Age, and Region are included in the Pre-COVID model with a cumulative weight of 0.0950. However, in the COVID model are completely absent; and in the Post-COVID and Global models, these borrower characteristics exhibit a marginal presence (weighting less than 0.0200).

The integration of C&R activities has yielded successful outcomes in our modelling exercise. These findings align with the work of several authors in the literature (such as Li et al., 2021; Bellotti et al., 2021; Han & Jang, 2013), who have confirmed that incorporating debt collection actions into LGD modelling leads to improved predictive results compared to models that rely solely on firm-specific variables.

These conclusions, which contrast with a significant portion of the literature, can primarily be attributed to two factors: the debt profile utilized in our research, and the widespread lack of access to information regarding C&R activities in existing studies. Our analysis is based on data from written-off portfolios with an average debt age of 11 years for Individuals and 12 years for SMEs, while most literature typically examines cases with much younger debt profiles.

In terms of RR modelling, borrower characteristics tend to be more relevant for newer debts, whereas for older debts, the importance of C&R activities increases, diminishing the relevance of borrower characteristics for predictions. Moreover, information concerning repayment behaviour and C&R activities is often challenging to access for researchers, limiting the incorporation of these variables into existing models. Fortunately, we had access to this data, which proved to be highly relevant for our modelling efforts.

In the context of C&R practices, the incorporation of macroeconomic variables into RR modelling is, comparatively, less prevalent. Nonetheless, our research indicates that these variables significantly enhance the predictive power of the constructed models. These variables are not only pertinent but also consistently featured among the final variables in all our modelling frameworks. However, as highlighted in previous literature (Kellner et al., 2022; Beck et al., 2017; Han & Jang, 2013), their relative weight within the models is diminished, exhibiting greater significance in the context of SMEs than in models pertaining to Individual borrowers.

The macroeconomic variables incorporated into our models align with widely recognized references in the academic literature, including the consumer price index, represented as GCPI and UCPI (Wu et al., 2024; Xu et al., 2024; Galow et al., 2024), GDP (Galow et al., 2024; Wang et al., 2020; Radivojević et al., 2019), the consumer confidence index (Xia et al., 2021), as well as activity and unemployment rates (Galow et al., 2024; Nazemi et al., 2022).

Focusing on the specifics of the Individual borrower models, we observe that the Judicial Flag emerges as one of the most critical variables across all constructed models. This variable indicates whether the account is currently involved in or has previously been subject to judicial claims post-acquisition by the Investment Firm. Its prominence is underscored by its weight in the Global model (0.5570) and the Pre-COVID model (0.4752); however, there is a notable decline associated with the impact of the COVID-19 pandemic, as evidenced by weights of 0.3230 in the COVID model and 0.1824 in the Post-COVID model.

The Judicial Flag's relevance is intrinsically linked to the C&R management strategies delineated by creditors, including both credit originators and investors, with respect to borrower profiles and debt characteristics that inform judicial actions. Typically, the criteria for targeting judicial activities involve several factors: the Debt Amount (with a minimum threshold established to account for fixed costs of judicial management), the Debt Age (usually a minimum of 90-120 days, allowing for amicable collection attempts prior), the Debt Product Type (wherein distinctions are made between secured and unsecured debts—given that judicial claims can vary significantly based on collateral), the Solvency of borrowers (where the presence of a positive solvency indicator is a common prerequisite for initiating judicial claims according to many creditor strategies), and the Trace/Untrace status of the debtors (as some institutions opt not to pursue judicial claims against debtors who have been previously untraceable).

Within the C&R management strategy employed by each entity, judicial claims are strategically defined to target accounts and debtors deemed more likely to fulfil their obligations, and the filing of lawsuits normally leads to an increase in recoveries. Our Judicial flag variable is inherently linked to a more favourable profile of debts and debtors. Despite its significance highlighted in our research, the presence of this variable in the existing modelling exercises is quite limited, with notable references only in Li et al. and Bailey (Li et al., 2022; Bailey, 2002). This scarcity is primarily due to restricted access to detailed C&R activities.

The observed decline in the importance of the Judicial flag in the models developed for the Pre-COVID, COVID, and Post-COVID periods can be attributed to the suspension of judicial activities prompted by the COVID pandemic. Government measures, such as movement restrictions and work-from-home policies, along with private entities' decisions to freeze certain C&R activities, have significantly contributed to substantial delays in judicial procedures.

Conversely, the Creditor variable, which encompasses the original creditor of the debt, emerges as another highly relevant factor in the Individual debtor models. Unlike the Judicial flag variable, its significance increased during the COVID period, ultimately becoming the most important variable in the Post-COVID model (with weights of 0.1709 for Pre-COVID, 0.3230 for COVID, and 0.5901 for Post-COVID). This trend is likely related to the reduced emphasis on the Judicial flag variable, stemming from the aforementioned halt in judicial activity. The existing literature reveals a limited examination of this variable (Nazemi et al., 2022), with predominant research focusing on debt arrangements from a single creditor. In contrast, our study broadens the scope by analysing the debt relationships involving multiple creditors.

The Judicial flag emerged as the only variable associated with C&R activities in the pre-COVID model. However, in the COVID and Post-COVID models, additional variables were introduced, specifically the Right-Party Contact (RPC) flag—which indicates whether the account experienced engagement with a right-party contact—and the Calls flag, which signifies whether the account received communication from the investor. Calls data have been utilized in the literature with positive outcomes, as evidenced by the work of Bellotti et al. and Ye & Bellotti, who identified them as highly relevant variables in their developed models (Bellotti et al., 2021; Ye & Bellotti, 2019). However, RPC is not commonly found in the literature.

In the COVID model, three iterations of the C&R activities were considered: the RPC flag for the last 90 days, the RPC ever flag, and the Calls ever flag. Conversely, the Post-COVID model incorporated the RPC flag for the last 90 days, the Calls flag for the last 90 days, along with the RPC ever and Calls ever flags. A notable shift can be observed in the COVID model, where the combined weights of the RPC and Calls flags totalled 0.3290, while in the Post-COVID model, this total diminished to 0.1808. This change in our models is likely directly attributable to the ramifications of the COVID-19 pandemic and the associated measures implemented therein. During the COVID period, all debtors were uniformly impacted by these measures, regardless of their risk profiles, thus amplifying the significance of C&R activities and their outcomes as reflected in our model weights.

The Debt Amount is recognized as a pivotal variable in the literature regarding LGD and RR modelling. Several scholars, including Xu et al., Li et al., and Jacobs, have identified the total debt amount as a significant variable within their developed models (Xu et al., 2024; Li et al., 2024; Jacobs, 2024). This finding underscores the importance of debt levels in understanding the dynamics addressed in their research. In the present study, this variable was accounted for through various representations of both original and current debt amounts, including total amount, principal amount, interest amount, and other amounts. Within our models for Individuals, the relevance of the Debt Amount was not pronounced. It played a significant role in the Pre-COVID model, with the initial weight of the other debt amount at 0.2476, positioning it as the second most significant variable, alongside a marginal weight of 0.0039 for the initial interest debt amount. However, its importance waned in the COVID model, where only the monthly debt amount was considered with a weight of 0.0239, leading to its omission in the Post-COVID model. This decline can be attributed to a combination of factors: firstly, the nature of the debt studied—comprising unsecured debts with an average age of 11 years—rendered the debt amount less significant. Secondly, the impact of COVID distinctly elevated the relevance of C&R activity variables, thereby diminishing the weight of variables associated with debt or borrower characteristics.

Product type has also been recognized as a relevant variable in existing literature (Nazemi et al., 2022; Kellner et al., 2022); however, its presence within our models for Individuals was marginal. It only appeared in the Post-COVID model with a limited weight of 0,0201. This outcome may be a result of our research being confined to unsecured debt and a portfolio characterized by an 11-year average debt age, wherein behavioural differences among product types are relatively minimal.

As previously mentioned, the significance of borrower characteristics in our Individuals models is relatively limited, in contrast with part of the literature (Li et al., 2024), largely due to the profile of the debt analysed in this research. The variables of Nationality, Age, and Region were included in the Pre-COVID model, collectively accounting for a total weight of 0.0950 (with Nationality at 0.0514, Age at 0.0314, and Region at 0.0122). None of these variables are present in the COVID model, and only Nationality appears marginally in the Post-COVID model with a weight of 0.0192. This slight shift in our modelling results due to the COVID effect can be attributed to the same reasons identified earlier regarding the increased weight of C&R activities variables, though in a contrasting direction.

These three variables have been noted in prior research, but generally, they do not carry significant weight in the constructed models. The ECB identifies Nationality as a variable to consider when segmenting NPLs (ECB, 2017). Although it is commonly acknowledged in C&R practice, its presence in academic literature as a variable of substantial importance in modelling is limited. In contrast, Age is much more prevalent in the literature, with numerous examples demonstrating its significant role in various developed models (Xu et al., 2024; Li et al., 2024; Bellotti et al., 2021). The Region, along with variations such as address, province, or area, is a commonly used determinant in the literature. We observe instances where it holds significant weight in models (Calabrese, 2012; Höcht & Zagst, 2010; Chen et al., 2009), while in other cases, it appears to have limited importance (Xu et al., 2024; Li et al., 2024; Kriebel & Yam, 2020).

Moreover, macroeconomic variables are included in all Individuals models, but they carry very marginal weight across the Pre-COVID, COVID, Post-COVID, and Global models. This aspect has remained unchanged despite the COVID impact. The marginal effect of these variables could be expected, as these variables are constant values for all the portfolios accounts, affecting equally to all of them and reducing its potential weight.

When examining the details of the SME models, we notice that the importance of variables has been less affected by COVID compared to the Individuals models. These SME models are highly concentrated around the top two variables: Guarantor flag and Judicial flag, which are the most relevant across all Pre-COVID, COVID, and Post-COVID models, with a combined weight ranging from 0.7000 to 0.8100.

The Guarantor flag variable serves a pivotal role in identifying accounts with associated guarantors and is regarded as the most significant variable in the developed models for

SMEs. Notably, the weight of this variable exhibited a slight increase during the COVID-19 pandemic; however, this increment lacked statistical significance (Pre-COVID weight: 0.4256; COVID weight: 0.4605; Post-COVID weight: 0.4763). In the context of C&R practices, this variable is extensively utilized for segmenting SME debt and for the formulation of targeted management strategies, both amicable and judicial. Entities with a guarantor typically demonstrate a higher probability of successful collections, particularly when the guarantor is an individual. The importance of this variable in our models substantiates this best practice within the C&R field.

In line with our research results and the C&R practice, the existing literature highlights a significant prevalence of variables associated with the presence of collateral or personal guarantees. Some studies emphasize the importance of factors such as the value, type, or proportion of collateral (Jacobs, 2024; Wang et al., 2020; Tanoue et al., 2017). Others point out the significance of having a guarantor involved (Li et al., 2022; Chen et al., 2009), while additional authors discuss the importance of both collateral and guarantors (Kellner et al., 2022; Höcht & Zagst, 2010; Dermine & De Carvalho, 2006).

The second most relevant variable within the SME models is the Judicial flag. Consistent with findings from our Individuals models, the weight of this variable decreased during the COVID-19 pandemic, albeit to a lesser extent (Pre-COVID weight: 0.3037; COVID weight: 0.3594; Post-COVID weight: 0.2249). This decline can be attributed to the suspension of judicial activities as a direct consequence of the pandemic and the related public and private measures enacted during this period. As noted in the discussion regarding the Individuals model, the Judicial flag also plays an important role, yet the existing literature on this variable is somewhat limited due to restricted access to comprehensive information on C&R activities. Nevertheless, references to the relevance of this information can be found in the works of Malinconico & Virguti, Li et al., and Bailey (Malinconico & Virguti, 2024; Li et al., 2022; Bailey, 2002).

Additionally, the Debt Amount variable exhibited a pattern analogous to that observed in the Individuals models. Initially, it held a significant weight in the Pre-COVID model (monthly debt amount ranked as the third most relevant variable, with a weight of 0.0895). The weight experienced a slight increase in the COVID model (Initial principal amount weight: 0.1068; Initial interest amount weight: 0.0044), yet became entirely negligible in the Post-COVID model. This phenomenon, similar to observations in the Individuals models, appears to arise from a combination of two factors: the specific debt profile analysed within this research (characterized by unsecured debts with an average duration of 12 years in SMEs) and the pandemic-induced elevation in the relevance of C&R activity-related variables.

The Product Type emerged as a significant variable within the models developed for SMEs, exhibiting a variation in weight across different time periods. In the Pre-COVID model, this variable possessed a weight of 0.0505, making it the fourth most relevant factor. However, it was entirely absent from the COVID model, likely reflecting the unprecedented market conditions during that period. In the Post-COVID model, its relevance diminished further, with a weight of 0.0236, positioning it as the fifth most significant variable. This observed decline in relevance may be attributed to the focus of the research on unsecured debt over a portfolio characterized by a 12-year debt age, wherein behavioural differences across product types become increasingly marginal.

As remarked in the discussion related to the Individuals modelling, Debt Amount is considered a prevalent determinant in the related literature (Xu et al., 2024; Li et al., 2024; Jacobs, 2024), and Product type has also been recognized as a relevant variable (Nazemi et al., 2022; Kellner et al., 2022).

In addition to the Judicial flag, several variables linked to C&R activities were incorporated into the SMEs models, although their presence varied significantly among the models. The Pre-COVID model included the Number of phone numbers (weight: 0.0446) and the Calls ever flag (weight: 0.0204), albeit with limited significance. The COVID model, in stark contrast, did not incorporate any additional C&R activity variables. The Post-COVID model, however, witnessed an increase in the relevance of such variables, as evidenced by the inclusion of the RPC ever flag (weight: 0.2000) and the Count of fixed-line phone numbers (weight: 0.0095). This shift in the models can largely be attributed to the transformative impact of the COVID-19 pandemic and the various public and private measures enacted, particularly the cessation of business activities among SMEs and the widespread implementation of work-from-home policies. These conditions rendered C&R activities less pertinent to debt recovery during the COVID period, a trend that markedly contrasts with the effects observed in the Individuals models. In the Post-COVID landscape, with the relaxation of previously implemented restrictions, the relevance of C&R activity variables experienced a resurgence.

The presence of contact information, such as mobile phone number, fixed-line phone number house and work phone numbers, and number of phone numbers available have been increasingly utilized in the literature with positive results (Kriebel & Yam, 2020; Ye & Bellotti, 2019; Shoghi, 2019). As well, the collections activities interactions, such as calls, visits, or SMS, among other, have been gaining relevant presence in the existing literature (Bellotti et al., 2021; Ye & Bellotti, 2019; Shoghi, 2019). However, due to the lack of access to detailed information of C&R activities, RPC is not commonly found in the literature.

Furthermore, the Creditor variable was included in all SMEs models, demonstrating a consistent yet limited weight across the different models: 0.0088 in the Pre-COVID model, 0.0187 in the COVID model, and 0.0166 in the Post-COVID model. This stability in weight suggests a relatively minor influence of the Creditor variable throughout the various phases of the economic landscape. This variable has not been extensively examined within the existing literature (Nazemi et al., 2022), as most research typically concentrates on debt from a singular creditor. In contrast, our study adopts a broader scope to explore this issue comprehensively.

Finally, the macroeconomic variables are incorporated into all the SME models with a reduced emphasis, appearing at similar levels across all constructed models, albeit higher than in the Individuals models. The variables included encompass various iterations of the Consumer Confidence Index (CCI), General Consumer Price Index (GCPI), Underlying Consumer Price Index (UCPI), Gross Domestic Product (GDP), and the Unemployment Rate (UR). These macroeconomic factors align with references in the existing literature, such as the consumer confidence index (Xia et al., 2021), the consumer price index (Wu et al., 2024; Xu et al., 2024), the GDP (Galow et al., 2024; Wang et al., 2020), and the unemployment rate (Galow et al., 2024; Nazemi et al., 2022).

5.4.- RECOMMENDATIONS BASED ON THE RESEARCH RESULTS

Besides its academic value, the findings from this research are likely to interest key stakeholders in the C&R sector, which can be categorized into two main groups: 1) creditors and servicers, and 2) regulators and supervisors.

For creditors and servicers managing debt portfolios similar to those examined in this study, there is a clear advantage in modelling potential RR in terms of both efficiency and effectiveness, and its application is for the debt portfolios management and for the valuation and pricing of debt portfolios to be sold or acquired. In the portfolios management, this approach enables these entities to concentrate their collection efforts on accounts that are more likely to yield successful recoveries, while minimizing efforts on accounts with lower collectability. And, in the selling and acquisition of portfolios, this approach facilitates a better understanding of the potential recoveries of the portfolio under analysis, improving the pricing capabilities and accuracy.

Additionally, incorporating debt characteristics, debtor traits, C&R activities, and macroeconomic factors as variables in the modelling process is beneficial. Specifically, debt characteristics and C&R activities emerge as the most pertinent sources of information for this modelling. We advise a particular focus on variables such as the Judicial flag, Creditor, and RPC in Individuals accounts, as well as the Judicial flag, Guarantor flag, and RPC for SMEs. Furthermore, it is advisable for these entities to refine their segmentation models for recovery rates in response to significant events like the COVID-19 pandemic, as these determinants have been shown to be influenced by such events, thereby increasing the importance of C&R activities and having a more pronounced effect on Individuals accounts compared to SMEs.

For regulators and supervisors, the insights derived from this research hold significance in several areas concerning the potential regulation and oversight of C&R activities. The segmentation models of RRs have demonstrated their effectiveness in enhancing both the efficacy and efficiency of C&R efforts, ultimately contributing to a stronger and more resilient economy. These models improve performance by considering a variety of variables, including debt characteristics, debtor profiles, C&R practices, and macroeconomic factors. Furthermore, events such as the COVID-19 pandemic (and the linked measures to alleviate its effects) have affected the primary determinants of recovery resources. Therefore, best practices in segmentation models should adapt to account for such occurrences in the future, which would bolster the resilience of creditor entities during economic downturns.

5.5.- LIMITATIONS AND STRENGTHS OF THE RESEARCH

This research has several limitations, notably its concentration on unsecured debts that are 11-12 years old. Including other types of debt, such as secured or more recent debts, could have enhanced the analysis. Furthermore, the absence of comprehensive data, such as details from judicial records, credit bureau information, scoring data, and financial insights from SMEs, hindered our modelling efforts.

In contrast, the developed models demonstrate significant strengths. They exhibit strong performance and align well with existing literature, highlighting their robustness, as the key variables remain stable over time, with fluctuations primarily occurring in their respective weights. These models were built on a comprehensive dataset comprising approximately 900.000 accounts monitored over a six-year period (2018-2023), which includes borrower characteristics, debt attributes, macroeconomic factors, and repayment behaviour, as well as C&R activities. Additionally, this study marks the first exploration of effective segmentation variables and the influence of COVID-19 on debt segmentation for the C&R of unsecured written-off portfolios acquired by investment firms in the Spanish market.

5.6.- FURTHER RESEARCH AREAS

Subsequent to the current investigation, several additional research avenues have been identified for future exploration: analyse the impact of collection results based on additional debt or debtor segments (like low, mid, and large-ticket debts); deliver the analysis and modelling exercise for a younger profile of debts; conduct a comparative analysis across different countries and regions; examine the transformation of business models within servicing and creditor entities in the collections and recoveries sector; assess the implications for specific debt sectors, including banking, financial institutions, telecommunications, utilities, and retail; evaluate the impact on various types of unsecured debt products such as loans, credit cards, and overdrafts; expand the research to include secured debt products that are not currently part of the study; analyse the stability of segmentation during stable cycles, avoiding disruptive events like those experienced during the COVID-19 pandemic; analyse the results by type of creditor (e.g., top-tier bank, mid-sized banks, saving banks, consumer credit entities); and analyse the results based on the borrowers' economic sector, as the measures and regulations implemented to mitigate COVID-19 effects varied considerably by activity sector.

SECTION 6.- REFERENCES

Alpaydin, E. (2021). *Machine learning*. MIT press.

Altman, E., Resti, A., & Sironi, A. (2004). Default recovery rates in credit risk modelling: a review of the literature and empirical evidence. *Economic notes*, *33*(2), 183-208.

Altman, E. I., Brady, B., Resti, A., & Sironi, A. (2005). The link between default and recovery rates: Theory, empirical evidence, and implications. *The Journal of Business*, *78*(6), 2203-2228.

Altman, E. I. (2021). Covid-19 and the credit cycle: 2020 revisited and 2021 outlook. *Journal of Credit Risk*.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance*, *23*(4), 589-609.

Asociación Nacional de Empresas de Gestión de Créditos. (2024). *Estudio de mercado sobre el sector de la recuperación de deuda y la morosidad*. ANGECO.

Athreya, K., Sánchez, J. M., Tam, X. S., & Young, E. R. (2018). Bankruptcy and delinquency in a model of unsecured debt. *International Economic Review*, *59*(2), 593-623.

Axis Corporate. (2025). *Assets Under Management Report. Edición VIII – 2025*. Axis Corporate, Accenture.

Baer, T. (2018). Behavioral insights and innovative treatments in collections. *McKinsey on Risk*, 5.

Bailey, M. (2002). *Consumer Collections & Recoveries: Operations and Strategies*. White Box.

Balzac, H. (1827). *El arte de pagar las deudas y satisfacer a los acreedores sin desembolsar un céntimo, en diez lecciones*. Editorial Centellas.

Basle Committee on Banking Supervision, & Bank for International Settlements. (2005). *An explanatory note on the Basel II IRB risk weight functions*. Bank for International Settlements.

Basle Committee on Banking Supervision, & Bank for International Settlements. (2025). *Principles for the management of credit risk*. Bank for International Settlements.

Basle Committee on Banking Supervision, & Bank for International Settlements. (2000). *Principles for the management of credit risk*. Bank for International Settlements.

Bastos, J. A. (2010). Forecasting bank loans loss-given-default. *Journal of Banking & Finance*, *34*(10), 2510-2517.

Bastos, J. A. (2014). Ensemble predictions of recovery rates. *Journal of Financial Services Research*, *46*(2), 177-193.

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

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

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

Bijak, K. (2014). Performance measures of models to predict Loss Given Default: a critical review.

Bijak, K., & Thomas, L. C. (2015). Modelling LGD for unsecured retail loans using Bayesian methods. *Journal of the Operational Research Society*, *66*(2), 342-352.

Brakes, M. (2020). The Debt Collection Outbreak: Recommendations for Idaho in Response to COVID-19. *Available at SSRN 3751680*.

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

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.

Calabrese, R., & Zanin, L. (2022). Modelling spatial dependence for Loss Given Default in peer-to-peer lending. *Expert Systems with Applications*, *192*, 116295.

Calabrese, R., & Zenga, M. (2010). Bank loan recovery rates: Measuring and nonparametric density estimation. *Journal of Banking & Finance*, *34*(5), 903-911.

Carleo, A., & Rocci, R. (2024). Functional clustering of NPLs recovery curves. *Socio-Economic Planning Sciences*, *95*, 102018.

Chalupka, R., & Kopecsni, J. (2008). *Modelling bank loan LGD of corporate and SME segments: a case study* (No. 27/2008). IES Working Paper.

Chen, H., Ma, Y.C., Chen, M. Z., Tang, Y., Wang, B., Chen, M., & Yang, X. G. (2009). Recovery Discrimination based on Optimized-Variables Support Vector Machine for Nonperforming Loan. *Systems Engineering-Theory & Practice*, 29(12), 23-30.

Cheng, D., & Cirillo, P. (2018). A reinforced urn process modeling of recovery rates and recovery times. *Journal of Banking & Finance*, *96*, 1-17.

Chin, A. G., & Kotak, H. (2006). Improving debt collection processes using rule-based decision engines: A case study of Capital One. *International Journal of Information Management*, 26(1), 81-88.

Coleman, A. M. (2004). *Collection management handbook: The art of getting paid*. John Wiley & Sons.

Cowling, M., Wilson, N., Nightingale, P., & Kacer, M. (2022). Predicting future default on the Covid-19 bounce back loan scheme: The £46.5 billion question. *International Small Business Journal*, *40*(5), 650-666.

Crespo, I., & Govindarajan, A. (2018). The analytics-enabled collections model. *McKinsey on payments. Special edition on Advanced Analytics in Banking*, *11*(30), 4-12.

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

Deville, J. (2014). Consumer credit default and collections: the shifting ontologies of market attachment. *Consumption Markets & Culture*, *17*(5), 468-490.

Drozd, L., & Serrano-Padial, R. (2014). *Modeling the revolving revolution: the role of IT reconsidered*. Wharton Working Paper.

Drozd, L. A., & Serrano-Padial, R. (2013). *Modeling the credit card revolution: the role of debt collection and informal bankruptcy* (No. 13-12). Federal Reserve Bank of Philadelphia.

Drozd, L. A., & Serrano-Padial, R. (2017). Modeling the revolving revolution: the debt collection channel. *American Economic Review*, *107*(3), 897-930.

Dumitrescu, E., Hué, S., Hurlin, C., & Tokpavi, S. (2022). Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *European Journal of Operational Research*, *297*(3), 1178-1192.

European Banking Authority. (2019). *EBA report on NPLs. Progress made and challenges ahead*. European Banking Authority.

European Central Bank. (2017). Guidance to banks on non-performing loans. *Guidance to Banks on Non-Performing Loans, Banking Supervision Division*, 1-131.

Fedaseyeu, V. (2015). *Debt collection agencies and the supply of consumer credit*. Research Department, Federal Reserve Bank of Philadelphia

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

Fedaseyeu, V., & Hunt, R. M. (2015). *The economics of debt collection: Enforcement of consumer credit contracts*. Research Department, Federal Reserve Bank of Philadelphia.

Fonseca, J., Strair, K., & Zafar, B. (2017). *Access to credit and financial health: evaluating the impact of debt collection* (No. 814, pp. 1-37). Federal Reserve Bank of New York.

Friedman, C. A., & Sandow, S. (2005). Estimating conditional probability distributions of recovery rates: a utility-based approach. Standard & Poor´s Risk Solutions Group.

Galow, B., Georgescu, O. M., & Marques, A. P. (2024). *Loss-given-default and macroeconomic conditions* (No. 2954). ECB Working Paper.

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

Giovine, C., Lerner, L., Thomas, R., Kakulavarapu, S., Chung, V., & Motwani, Y. (2024). Extracting value from AI in banking: Rewiring the enterprise. *McKinsey & Company*.

Grunert, J., & Weber, M. (2005). Recovery rates of bank loans: empirical evidence for Germany.

Gupton, G. M., Gates, D., & Carty, L. V. (2000). Bank loan loss given default. *Moody's Investors Service, Global Credit Research, November*.

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

Han, C., & Jang, Y. (2013). Effects of debt collection practices on loss given default. *Journal of Banking & Finance*, *37*(1), 21-31.

Higginson, M., Jacques, F., & Rudisuli, R. (2018). The seven pillars of (collections) wisdom. *McKinsey on Risk*, *5*.

Höcht, S., & Zagst, R. (2010). Loan Recovery Determinants--A Pan-European Study. *Available at SSRN 2443724*.

Hoechstoetter, M., Nazemi, A., Rachev, S. T., & Bozic, C. (2012). Recovery rate modelling of non-performing consumer credit using data mining algorithms. *Review of*.

Hoechstoetter, M., & Nazemi, A. (2013). Analysis of loss given default. *Investment management and financial innovations*, (10, Iss. 4), 70-79.

Hunt, R. M. (2007). Collecting consumer debt in America. *Available at SSRN 993249*.

Informa, D.B.K. (2023). Empresas de gestión de impagados. Informes Especiales.

Jacobs, M. (2024). Modeling Ultimate Loss-Given-Default and Time-to-Resolution on Corporate Debt. *Available at SSRN 4738268*.

Jimenez, G., & Saurina, J. (2006). Credit cycles, credit risk, and prudential regulation. *International Journal of Central Banking*, 2(2).

Jonasson, T., Papaioannou, M. G., & Williams, M. (2019). Debt management. *Sovereign Debt: A Guide for Economists and Practitioners*, *192*.

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

Khieu, H. D., Mullineaux, D. J., & Yi, H. C. (2012). The determinants of bank loan recovery rates. *Journal of Banking & Finance*, *36*(4), 923-933.

Kriebel, J., & Yam, K. (2020). Forecasting recoveries in debt collection: Debt collectors and information production. *European Financial Management*, *26*(3), 537-559.

Krumbein, W. C. (1924). Collection Agencies. *University Journal of Business*, 48-67.

Laitinen, E. K. (2025). Gini Coefficient and AUC in Assessing Predictive Model Performance: Effect of Ranks. *Theoretical Economics Letters, 15,* 1251-1266.

Laurent, M. P., Plantefeve, O., Tejada, M., & Van Weyenbergh, F. (2020). Banking models after COVID-19: Taking model-risk management to the next level. *McKinsey & Company. Recovered from https://www.mckinsey.com/business-functions/risk/our-insights/bankingmodels-after-covid-19-taking-model-risk-management-to-the-next-level*.

Lawrence, D. B. (1992). *Handbook of consumer lending*. Englewood Cliffs, NJ, Prentice-Hall.

Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European journal of operational research, 247*(1), 124-136.

Li, A., Li, Z., & Bellotti, A. (2023A). Predicting loss given default of unsecured consumer loans with time-varying survival scores. *Pacific-Basin Finance Journal*, *78*, 101949.

Li, G., Chen, M., & Yang, X. (2022, July). Loss Given Default Modeling and Variable Importance Analysis of Credit Loans. In *2022 41st Chinese control conference (CCC)* (pp. 7491-7496). IEEE.

Li, J., Kang, J., Wu, J., Wang, H., & Yang, X. (2024). Research on credit card default repayment prediction model. *The Journal of Finance and Data Science*, *10*, 100136.

Li, K., Zhou, F., Li, Z., Yao, X., & Zhang, Y. (2021). Predicting loss given default using post-default information. *Knowledge-Based Systems*, *224*, 107068.

Li, Z., Li, A., Bellotti, A., & Yao, X. (2023B). The profitability of online loans: A competing risks analysis on default and prepayment. *European Journal of Operational Research*, *306*(2), 968-985.

Lin, Z. Y., & Lu, J. W. (2024). The impact of COVID-19 on global investor attention. *Pacific-Basin Finance Journal*, *88*, 102522.

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

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

Malikkidou, D., & Strohbach, W. (2025). Predicting Bank Distress in Europe: Using Machine Learning and a Novel Definition of Distress. *European Banking Authority Research Paper*, (21).

Malinconico, A., & Virguti, E. (2024). Business Models and Best Practices in Debt Recovery Services: an overview of the Italian industry. In *Conference Proceedings*.

Matuszyk, A., Mues, C., & Thomas, L. C. (2010). Modelling LGD for unsecured personal loans: Decision tree approach. *Journal of the Operational Research Society*, *61*(3), 393-398.

Molina, S. (2015). *El ciclo del crédito*. Editorial Almuzara.

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

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

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

Nembrini, S., König, I. R., & Wright, M. N. (2018). The revival of the Gini importance?. *Bioinformatics, 34*(21), 3711-3718.

O'Brien, L., Chen, V., Ramsay, I., & Ali, P. (2021). An Impending 'Avalanche': Debt Collection and Consumer Harm after COVID-19. *Australian Business Law Review*, *49*(2), 84-115.

Piumelli, F. A., & Schmidt, D. A. (1998). *Power collecting: automation for effective asset management*. John Wiley & Sons.

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

Radivojević, N., Cvijanović, D., Sekulic, D., Pavlovic, D., Jovic, S., & Maksimović, G. (2019). Econometric model of non-performing loans determinants. *Physica A: Statistical Mechanics and its Applications*, *520*, 481-488.

Ramos González, M. R., Ureña, A. P., & Fernández-Aguado, P. G. (2023). Forecasting for regulatory credit loss derived from the COVID-19 pandemic: A machine learning approach. *Research in International Business and Finance*, *64*, 101907.

Repullo, R., Saurina, J., & Trucharte, C. (2010). Mitigating the Pro-cyclicality of Basel II. *Economic Policy*, *25*(64), 659-702.

Rosenberg, E., & Gleit, A. (1994). Quantitative methods in credit management: a survey. *Operations research*, *42*(4), 589-613.

SAS (2025). Transforming Risk Management. SAS & FT Longitude

Schuermann, T. (2004). What do we know about loss given default?

Serrano-Cinca, C., Gutiérrez-Nieto, B., & López-Palacios, L. (2015). Determinants of default in P2P lending. *PloS one*, *10*(10), e0139427.

Shoghi, A. (2019). Debt collection industry: machine learning approach. *Journal of Money and Economy*, *14*(4), 453-473.

Sopitpongstorn, N., Silvapulle, P., Gao, J., & Fenech, J. P. (2021). Local logit regression for loan recovery rate. *Journal of Banking & Finance*, *126*, 106093.

Spuchľakova, E., & Cug, J. (2015). Credit risk and LGD modelling. *Procedia Economics and Finance*, *23*, 439-444.

Stifler, L., & Parrish, L. (2014). *Debt Collection & Debt buying. The State of Lending in America & its Impact on the U.S. Households*. Center for Responsible Lending.

Tanoue, Y., Kawada, A., & Yamashita, S. (2017). Forecasting loss given default of bank loans with multi-stage model. *International Journal of Forecasting*, *33*(2), 513-522.

Thomas, L., & Bijak, K. (2015). *Impact of segmentation on the performance measures of LGD models*.

Thomas, L. C., Matuszyk, A., So, M. C., Mues, C., & Moore, A. (2016). Modelling repayment patterns in the collections process for unsecured consumer debt: A case study. *European Journal of Operational Research*, *249*(2), 476-486.

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

Thomas, L. C., Matuszyk, A., & Moore, A. (2009). Collections policy comparison in LGD modelling.

Várgedő, B. (2024). A stochastic loss given default model of Hungarian residential mortgages. *Acta Oeconomica*, *74*(3), 299-327.

Wang, H., Forbes, C. S., Fenech, J. P., & Vaz, J. (2020). The determinants of bank loan recovery rates in good times and bad–new evidence. *Journal of Economic Behavior & Organization*, *177*, 875-897.

Wu, Y., Arora, G., & Mei, X. (2024). Using CPI in Loss Given Default Forecasting Models for Commercial Real Estate Portfolio. *arXiv preprint arXiv:2402.15498*.

Xia, Y., Zhao, J., He, L., Li, Y., & Yang, X. (2021). Forecasting loss given default for peer-to-peer loans via heterogeneous stacking ensemble approach. *International Journal of Forecasting*, *37*(4), 1590-1613.

Xu, Q., Liu, C., Luo, J., & Liu, F. (2024). Using machine learning to investigate the determinants of loan default in P2P lending: Are there differences between before and during COVID-19?. *Pacific-Basin Finance Journal*, *88*, 102550.

Ye, H., & Bellotti, A. (2019). Modelling recovery rates for non-performing loans. *Risks*, *7*(1), 19.

Zhang, J., & Thomas, L. C. (2012). Comparisons of linear regression and survival analysis using single and mixture distributions approaches in modelling LGD. *International Journal of Forecasting*, *28*(1), 204-215.

SECTION 7.- APPENDICES

APPENDIX 1.- DATA FILES DETAIL

APPENDIX 2.- DATA CLEANING AND NORMALIZATION DETAIL

APPENDIX 3.- EXPLANATORY VARIABLES DETAIL

APPENDIX 4.- MACROECONOMIC FACTORS DETAIL

APPENDIX 5.- MODELS DETAIL

APPENDIX 6.- THESIS SUMMARY

APPENDIX 1.- DATA FILES DETAIL

1- File "CarterasFechaInicio": file with information about the portfolio acquisition by the Investment Fund and the debt vertical to which it belongs.

a. File fields

Table 99. Fields in the File "CarterasFechaInicio"

Field	Description
"Cartera"	Portfolio: name of the debt portfolio acquired
"Fecha_alta"	Date of acquisition: date of the portfolio acquisition by the Investment Fund.
"Vertical"	Debt vertical: client segment/typology

Source: Own elaboration

b. File extract example

Figure 92. File "CarterasFechaInicio" extract example

Cartera	Fecha_alta	Vertical
A	20/8/2015	Individual
A	20/8/2015	SME
A	20/8/2015	Individual
A	20/8/2015	SME
A	16/4/2019	Individual
A	22/10/2019	Secured
A	17/2/2022	Individual
A	17/2/2022	Individual
B	17/2/2022	Individual
B	17/2/2022	Individual

Source: Own elaboration

2- File "Cuenta": file with information about the accounts and original and actual debt from the Individuals segment.

a. File fields

Table 100. Fields in the File "Cuenta"

Field	Description
"ID_CUENTA"	Account ID: account identifier
"MARCA_INDIVIDUAL_EMPRESA"	Flag Individual-SME: clients segment / typology
"CARTERA"	Portfolio: name of the debt portfolio acquired
"SERIE"	Series: account identifier

“NUMERO”	Number: account identifier
“IUC”	Account ID: account identifier
“DEUDA_INICIAL”	Initial debt: client´s initial debt originated
“DEUDA_ACTUAL”	Actual debt: client´s actual debt
“DEUDA_ACTUAL_INTERESES”	Actual interest debt: interests generated by the debt
“DEUDA_ACTUAL_NOMINAL”	Actual nominal debt: nominal amount of the debt
“DEUDA_ACTUAL_OTROS”	Actual others debt: other costs originated by the debt
“DEUDA_INICIAL_INTERESES”	Initial interest debt: interests generated by the debt initially
“DEUDA_INICIAL_NOMINAL”	Initial nominal debt: nominal amount of the debt initially
“DEUDA_INICIAL_OTROS”	Initial others debt: other costs originated by the debt initially
“ESTADO_PROC”	Procedure status: status of the procedure
“FECHA_CANCELACION”	Cancellation date: date of the debt cancellation
“FECHA_IMPAGO”	Default date: date of the debt default
“CEDENTE”	Debt seller: originator and seller of the debt acquired
“TIPO_PRODUCTO”	Debt product type: type of product of the debt originated

Source: Own elaboration

b. File extract example

Figure 93. File “Cuenta” extract example

ID_CUENTA	MARCA_INDIVIDUAL_EMPRESA	CARTERA	SERIE	NUMERO	IUC	DEUDA_INICIAL	DEUDA_ACTUAL	DEUDA_ACTUAL_INTERESES	DEUDA_ACTUAL_NOMINAL	DEUDA_ACTUAL_OTROS	DEUDA_INICIAL_INTERESES
51037	I	W	8311	985	1870454	2875,47	2875,47	0	2875,47	0	0
51038	I	W	8311	3030	2818697	2698,67	-8540,71	0	0	-8540,71	0
51039	I	W	8311	987	1875759	6991,9	146,73	0	262,16	-115,43	0
51040	I	W	8311	988	2168277	12302,36	12302,36	0	12302,36	0	0
51041	I	W	8311	989	1870455	4734,24	4706,28	0	4706,28	0	0
51042	I	W	8311	990	1870456	5954,56	4427,49	0	4427,49	0	0
51043	I	W	8311	991	1875760	3851,63	-1246,01	0	-1246,01	0	0
51044	I	W	8311	992	1870457	4173,76	4173,76	0	4173,76	0	0
51045	I	W	8311	993	1870458	372,11	0	0	0	0	0
51046	I	W	8311	994	1875761	3984,31	3984,31	0	3984,31	0	0

DEUDA_INICIAL_NOMINAL	DEUDA_INICIAL_OTROS	ESTADO_PROC	FECHA_CANCELACION	FECHA_IMPAGO	CEDENTE	TIPO_PRODUCTO
2875,47	0	V		15/10/1998	CAJASTUR	VARIOS
2698,67	0	B	22/11/2019	21/10/2004	CAJASTUR	VARIOS
6991,9	0	B	12/5/2022	9/9/1999	CAJASTUR	VARIOS
12302,36	0	V		20/10/2005	CAJASTUR	VARIOS
4734,24	0	V		7/3/2003	CAJAEXTRAMADURA	VARIOS
5954,56	0	V		18/10/2007	CAJASTUR	VARIOS
3851,63	0	B	8/3/2022	22/10/2001	CAJASTUR	VARIOS
4173,76	0	V		1/7/1992	CAJASTUR	VARIOS
372,11	0	V		8/2/2011	CCM	VARIOS
3984,31	0	V		14/7/2000	CAJASTUR	VARIOS

Source: Own elaboration

3- **File “Cuentas_SME”**: file with information about the accounts and original and actual debt from the SMEs segment.

a. File fields

Table 101. Fields in the File “Cuentas_SME”

Field	Description
“ID_CUENTA”	Account ID: account identifier
“MARCA_INDIVIDUAL_EMPRESA”	Flag Individual-SME: clients segment / typology
“CARTERA”	Portfolio: name of the debt portfolio acquired
“SERIE”	Series: account identifier
“NUMERO”	Number: account identifier
“IUC”	Account ID: account identifier
“DEUDA_INICIAL”	Initial debt: client´s initial debt originated
“DEUDA_ACTUAL”	Actual debt: client´s actual debt
“DEUDA_ACTUAL_INTERESES”	Actual interest debt: interests generated by the debt
“DEUDA_ACTUAL_NOMINAL”	Actual nominal debt: nominal amount of the debt
“DEUDA_ACTUAL_OTROS”	Actual others debt: other costs originated by the debt
“DEUDA_INICIAL_INTERESES”	Initial interest debt: interests generated by the debt initially
“DEUDA_INICIAL_NOMINAL”	Initial nominal debt: nominal amount of the debt initially
“DEUDA_INICIAL_OTROS”	Initial others debt: other costs originated by the debt initially
“ESTADO_PROC”	Procedure status: status of the procedure
“FECHA_CANCELACION”	Cancellation date: date of the debt cancellation
“FECHA_IMPAGO”	Default date: date of the debt default
“CEDENTE”	Debt seller: originator and seller of the debt acquired
“TIPO_PRODUCTO”	Debt product type: type of product of the debt originated

Source: Own elaboration

b. File extract example

Figure 94. File “Cuentas_SME” extract example

ID_CUENTA	MARCA_INDIVIDUAL_EMPRESA	CARTERA	SERIE	NUMERO	IUC	DEUDA_INICIAL	DEUDA_ACTUAL	DEUDA_ACTUAL_INTERESES	DEUDA_ACTUAL_NOMINAL	DEUDA_ACTUAL_OTROS	DEUDA_INICIAL_INTERESES
53120	E	B	8392	146	2036125	450,19	450,19	0	450,19	0	0
53121	E	B	8392	147	2036126	2781,13	2781,13	0	2781,13	0	0
53122	E	B	8392	148	1922559	23462,99	23462,99	0	23462,99	0	0
53123	E	B	8392	149	2036127	2809,03	2759,64	0	2759,64	0	0
53124	E	B	8392	150	2435027	139,49	139,49	0	139,49	0	0
53125	E	B	8392	151	2036128	2402,64	2402,64	0	2402,64	0	0
53126	E	B	8392	152	2036129	329,32	329,32	0	329,32	0	0
53127	E	B	8392	153	2036130	3812,54	3812,54	0	3812,54	0	0
53128	E	B	8392	154	1904738	81529,8	81515,6	0	81515,6	0	0
53129	E	B	8392	155	2036131	501,59	501,59	0	501,59	0	0

DEUDA_INICIAL_NOMINAL	DEUDA_INICIAL_OTROS	ESTADO_PROC	FECHA_CANCELACION	FECHA_IMPAGO	CEDENTE	TIPO_PRODUCTO
450,19	0	V		25/3/2010		VARIOS
2781,13	0	V		3/2/2009		VARIOS
23462,99	0	V		6/11/2008		VARIOS
2809,03	0	V		2/2/2011		VARIOS
139,49	0	V		12/1/2009		VARIOS
2402,64	0	V		29/4/2009		VARIOS
329,32	0	V		8/4/2009		VARIOS
3812,54	0	V		2/6/2009		VARIOS
81529,8	0	V		25/11/2009		VARIOS
501,59	0	V		21/12/2010		VARIOS

Source: Own elaboration

4- File “Cuentas_Gesco00-06”: file with information about the accounts and original and actual debt from both Individuals and SMEs segments.

a. File fields

Table 102. Fields in the File “Cuentas_Gesco00-06”

Field	Description
“EXP_SERIE”	Series: account internal identifier
“EXP_NUMERO”	Number: account identifier
“CLI_NEGOCIO”	Vertical flag: vertical identifier
“CLI_COMPRADOR_CARTERA”	Portfolio name: name of the portfolio acquired
“EXPPCO_ID_PRESENCE”	Account ID: account identifier
“DEUDA_INICIAL”	Initial debt: client´s initial debt originated
“DEUDA_ACTUAL”	Actual debt: client´s actual debt
“DEUDA_ACTUAL_INTERESES”	Actual interest debt: interests generated by the debt
“DEUDA_ACTUAL_NOMINAL”	Actual nominal debt: nominal amount of the debt
“DEUDA_ACTUAL_OTROS”	Actual others debt: other costs originated by the debt
“DEUDA_INICIAL_INTERESES”	Initial interest debt: interests generated by the debt initially
“DEUDA_INICIAL_NOMINAL”	Initial nominal debt: nominal amount of the debt initially
“DEUDA_INICIAL_OTROS”	Initial others debt: other costs originated by the debt initially
“EXP_SITUACION”	Procedure status: status of the procedure
“EXP_FECHA_CANCELACION”	Cancellation date: date of the debt cancellation
“FECHA_PRIMER_IMPAGO”	First default date: date of the first debt default
“ENTITY”	Debt seller: originator and seller of the debt acquired
“TIPO_PRODUCTO”	Debt product type: type of product of the debt originated

Source: Own elaboration

b. File extract example

Figure 95. File "Cuentas_Gesco00-06" extract example

EXP_SERIE	EXP_NUMERO	CLI_NEGOCIO	CLI_COMPRADOR_CARTERA	EXPPCO_ID_PRESENCE	DEUDA_INICIAL	DEUDA_ACTUAL	DEUDA_ACTUAL_INTERESES	DEUDA_ACTUAL_NOMINAL	DEUDA_ACTUAL_OTROS	DEUDA_INICIAL_INTERESES
1512	856	S	M	22493341	182,42	182,42	0	182,42	0	0
1533	7619	I	A	22517323	86,64	86,64	0	86,64	0	0
3089	18293	I	E	19586623	2011,14	2011,14	572,34	1438,8	0	572,34
5811	8486	I	M	2094805	4055,2	3544,36	1647,68	1896,68	0	1647,68
8153	6174	I	S	14143546	2206,33	1263,89	0	1263,89	0	0
3069	292818	I	A	14621404	708,7	708,7	0	708,7	0	0
3069	532266	I	A	14860829	317,83	317,83	0	317,83	0	0
7163	20578	I	V	24066936	1964,03	489,03	96,68	313,4	78,95	96,68
3069	371525	I	A	14700101	209,33	209,33	0	209,33	0	0
5013	40574	I	M	2079984	73,32	73,32	0	73,32	0	0

DEUDA_INICIAL_NOMINAL	DEUDA_INICIAL_OTROS	EXP_SITUACION	EXP_FECHA_CANCELACION	FECHA_PRIMER_IMPAGO	ENTITY	TIPO_PRODUCTO
182,42	0	A	NULL	2005-01-04 00:00:00.000	NULL	OVERDRAFTS
86,64	0	A	NULL	2018-10-01 00:00:00.000	NULL	CREDIT CARD
1438,8	0	A	NULL	2011-03-01 00:00:00.000	UNICAJA	CRÉDITO TARJETA
2407,52	0	A	NULL	2011-08-01 00:00:00.000	NULL	Préstamo
2206,33	0	A	NULL	2013-04-11 00:00:00.000	Banco Santander	Credit Card
708,7	0	A	NULL	2011-10-21 00:00:00.000	ORANGE	ORANGE MOVIL
317,83	0	A	NULL	2011-05-09 00:00:00.000	JAZZTEL	Fijo y Móvil Jazztel
1788,4	78,95	C	2023-01-30 00:00:00.000	2020-01-03 00:00:00.000	NULL	LOAN
209,33	0	A	NULL	2013-02-14 00:00:00.000	ORANGE	ORANGE MOVIL
73,32	0	A	NULL	2012-12-01 00:00:00.000	CaixaBank	Tarjeta de Crédito

Source: Own elaboration

5- **File "Persona"**: file with information about the person who originated the debt.

a. File fields

Table 103. Fields in the File "Persona"

Field	Description
"ID_CUENTA"	Account ID: account identifier
"ID_PERSONA"	Person ID: person identifier
"TIPO_PARTICIPANTE"	Participant type: type of participant in the debt
"FECHA_NACIMIENTO"	Date of birth: date of birth of the person
"TIPO_DOC"	Document type: type of document of the person
"NUM_TEL_MOVIL"	Number of mobile numbers: number of mobile numbers of the person
"NUM_TEL_FIJO"	Number of fixed line phone numbers: number of fixed line phone numbers of the person

Source: Own elaboration

b. File extract example

Figure 96. File "Persona" extract example

ID_CUENTA	ID_PERSONA	TIPO_PARTICIPANTE	FECHA_NACIMIENTO	TIPO_DOC	NUM_TEL_MOVIL	NUM_TEL_FIJO
51947	75037	TIT		DESCONOCIDO	3	
61711	75037	TIT		DESCONOCIDO	3	
64722	75037	TIT		DESCONOCIDO	3	
52100	75038	TIT	22/8/1956	NIF	1	1
333991	75038	TIT	22/8/1956	NIF	1	1
52100	75039	TIT	28/3/1947	NIF	3	1
333991	75039	TIT	28/3/1947	NIF	3	1
52101	75040	TIT	6/4/1934	NIF	2	1
52101	75041	TIT	15/5/1934	NIF	2	1
52102	75042	TIT	31/1/1967	NIF	4	2

Source: Own elaboration

6- File “Persona_Gesco”: file with information about the person which originated the debt.

a. File fields

Table 104. Fields in the File “Persona_Gesco”

Field	Description
“EXP_SERIE”	Series: account internal identifier
“EXP_NUMERO”	Number: account identifier
“PER_NUMERO”	Number: person identifier
“TIPO_PARTICIPANTE”	Participant type: type of participant in the debt
“FECHA_NACIMIENTO”	Date of birth: date of birth of the person
“TIPO_DOC”	Document type: type of document of the person
“NUM_TEL_MOVIL”	Number of mobile numbers: number of mobile numbers of the person
“NUM_TEL_FIJO”	Number of fixed line phone numbers: number of fixed line phone numbers of the person

Source: Own elaboration

b. File extract example

Figure 97. File “Persona_Gesco” extract example

EXP_SERIE	EXP_NUMERO	PER_NUMERO	TIPO_PARTICIPANTE	FECHA_NACIMIENTO	TIPO_DOC	NUM_TEL_MOVIL	NUM_TEL_FIJO
7312	773	6253007	BLOQUEADO AML	29/3/1999	CIF		
3177	7923	9293117	INDETER				
3032	2025	8294309	DEUDOR		CIF		1
3084	54	5313506	VINCULACION	25/8/1978	NIF	1	3
3073	524	5140806	DEUDOR	1/3/1986	NIF	1	
3177	15772	9297224	INDETER				
3097	1900	5381572	DEUDOR	2/6/1955	NIF	1	3
7312	127	5314882	DEUDOR	22/3/1953	NIF		
8199	5279	475101	DEUDOR	27/12/1953	NIF	1	
3094	353	5377825	DEUDOR	26/9/1987			

Source: Own elaboration

7- File “Codigo_Postal”: file with information about the accounts postal code.

a. File fields

Table 105. Fields in the File “Codigo_Postal”

Field	Description
“ID_PERSONA”	Person ID: person identifier
“CODIGO_POSTAL”	Postal code: postal code of the person
“ORDINALIDAD”	Ordinality: ordinality in case of having several postal codes
“FECHA_ALTA”	Register date: date of the postal code registry

Source: Own elaboration

b. File extract example

Figure 98. File "Codigo_Postal" extract example

ID_PERSONA	CODIGO_POSTAL	ORDINALIDAD	FECHA_ALTA
2	28330	0	6/2/2008
2	28224	2	6/2/2008
3	28022	0	7/4/2011
3	28330	0	1/6/2015
4	19170	0	16/6/2010
4	28108	3	27/6/2016
4	28703	6	4/7/2013
5	28703	0	1/6/2015
6	7014	0	4/11/2009
6	7013	0	1/6/2015

Source: Own elaboration

8- File "CP_Gesco": file with information about the accounts postal code.

a. File fields

Table 106. Fields in the File "CP_Gesco"

Field	Description
"PER_NUMERO"	Person ID: person identifier
"Per_CP"	Postal code: person postal code

Source: Own elaboration

b. File extract example

Figure 99. File "CP_Gesco" extract example

PER_NUMERO	per_cp
7	29600
1554300	29712
62962	8014
39082	3690
39299	25005
41889	43006
49945	8130
52632	8221
1561803	13160
1550842	28806

Source: Own elaboration

9- File "Procedimiento": file with judicial information of the accounts.

a. File fields

Table 107. Fields in the File "Procedimiento"

Field	Description
"ID_CUENTA"	Account ID: account identifier
"IUC"	Account ID: account identifier
"MARCA_INDIVIDUAL_EMPRESA"	Flag Individual-SME: clients segment / typology

"CARTERA"	Portfolio: name of the debt portfolio acquired
"ID_PROCEDI"	Procedure ID: judicial procedure identifier
"ESTADO_PROC"	Procedure status: status of the judicial procedure
"FECHA_DEMANDA"	Legal claim date: date of the legal claim

Source: Own elaboration

b. File extract example

Figure 100. File "Procedimiento" extract example

ID_CUENTA	IUC	MARCA_INDIVIDUAL_EMPRESA	CARTERA	ID_PROCEDI	ESTADO_PROC	FECHA_DEMANDA
53120	2036125	E	B	2	V	
53121	2036126	E	B	3	V	
53122	1922559	E	B	4	V	
53123	2036127	E	B	5	V	
53124	2435027	E	B	6	V	
53125	2036128	E	B	4299	V	
53126	2036129	E	B	8	V	
53127	2036130	E	B	9	V	
53128	1904738	E	B	10	V	
53129	2036131	E	B	11	V	

Source: Own elaboration

10- File "Procedimiento2": file with judicial information of the accounts.

a. File fields

Table 108. Fields in the File "Procedimiento2"

Field	Description
"ID_CUENTA"	Account ID: account identifier
"IUC"	Account ID: account identifier
"MARCA_INDIVIDUAL_EMPRESA"	Flag Individual-SME: clients segment / typology
"CARTERA"	Portfolio: name of the debt portfolio acquired
"ID_PROCEDI"	Procedure ID: judicial procedure identifier
"ESTADO_PROC"	Procedure status: status of the judicial procedure
"NUMERO_AUTOS"	Court order number: number of the procedure court order
"FECHA_BAJA"	Withdrawal date: date of the procedure withdrawal
"FECHA_FIN"	End date: date of the end of the procedure

Source: Own elaboration

b. File extract example

Figure 101. File "Procedimiento2" extract example

ID_CUENTA	IUC	MARCA_INDIVIDUAL_EMPRESA	CARTERA	ID_PROCEDI	ESTADO_PROC	NUMERO_AUTOS	FECHA_BAJA	FECHA_FIN
53120	2036125	E	B	2	V	2011		
53121	2036126	E	B	3	V	2009		
53122	1922559	E	B	4	V	2009		
53123	2036127	E	B	5	V	2011		
53124	2435027	E	B	6	V	2009		
53126	2036129	E	B	8	V	2009		
53127	2036130	E	B	9	V	2010		
53128	1904738	E	B	10	V	2010		
53129	2036131	E	B	11	V	2011		
53130	2064693	E	B	12	V	2009		

Source: Own elaboration

11- File "SMS": file with information of the SMSs sent.

a. File fields

Table 109. Fields in the File "SMS"

Field	Description
"ID_CUENTA"	Account ID: account identifier
"CARTERA"	Portfolio: name of the debt portfolio acquired
"MARCA_INDIVIDUAL_EMPRESA"	Flag Individual-SME: clients segment / typology
"ID"	SMS ID: SMS identifier
"FECHA_SMS"	SMS date: date of the SMS sending
"MODELO_COMUNICADO"	Release model: type and model of the SMS sent

Source: Own elaboration

b. File extract example

Figure 102. File "SMS" extract example

ID_CUENTA	CARTERA	MARCA_INDIVIDUAL_EMPRESA	ID	FECHA_SMS	MODELO_COMUNICADO
277006	I	I	82	25/5/2020	SMS DEMJ
47941	I	I	109	29/5/2020	SMS NUM 1B
275103	I	I	130	29/5/2020	SMS QUITA
304002	W	I	146	30/5/2020	PROM4
275247	I	I	173	1/6/2020	SMS NUM 1B
304002	W	I	176	1/6/2020	PROM4
276741	I	I	192	1/6/2020	PROM2
276248	I	I	132	29/5/2020	SMS QUITA
288744	I	I	169	1/6/2020	PROM2
276899	I	I	362	5/6/2020	PROM2

Source: Own elaboration

12- File "Total_Llamadas": file with information of the calls and actions delivered.

a. File fields

Table 110. Fields in the File "Total_Llamadas"

Field	Description
"YEAR"	Year: year of the action

“MONTH”	Month: month of the action
“N_LLAMADAS”	Number of calls: number of calls
“RPC”	Right party contact: identifier of right party contact
“PTP”	Promise to pay: identifier of promise to pay
“IUC”	ID: account identifier
“ID_CUENTA”	Account ID: account identifier
“CARTERA”	Portfolio: name of the debt portfolio acquired
“SERIE”	Series: account identifier
“NUMERO”	Number: account identifier
“ID_PRESENCE”	Presence ID: action identifier

Source: Own elaboration

b. File extract example

Figure 103. File “Total_Llamadas” extract example

YEAR	MONTH	N_LLAMADAS	RPC	PTP	IUC	ID_CUENTA	CARTERA	SERIE	NUMERO	ID_PRESENCE
2021	5	28	0	0	17131095	553222	C	3019	1805	17131095
2020	11	12	2	0	17131471	553660	C	3019	2289	17131471
2022	1	12	0	0	17003614	584200	S	3029	28560	17003614
2021	7	3	0	0	16976022	595100	S	3029	40505	16976022
2021	2	5	0	0	14095508	530011	S	3053	414	14095508
2020	10	2	0	0	14103627	536762	S	3053	8535	14103627
2021	9	2	0	0	1649324	280550	I	3081	3787	1649324
2022	4	3	0	0	1649666	61354	I	3081	5702	1649666
2022	2	3	0	0	3372953	286217	I	3081	10331	3372953
2020	10	1	0	0	1778451	287505	I	3081	11749	1778451

Source: Own elaboration

13- File “Cobros”: file with information of the account’s payments.

a. File fields

Table 111. Fields in the File “Cobros”

Field	Description
“COB_NUMERO”	Payment ID: identifier of the payment
“IUC”	ID: account identifier
“SERIE”	Serie: account identifier
“NUMERO”	Number: account identifier
“Cartera”	Portfolio: name of the debt portfolio acquired
“Cli_negocio”	Client type (vertical flag): type of client (vertical identifier)
“Collection_date”	Collection date: date of the payment
“IMPORTE”	Amount: payment amount
“INTERIM”	Interim flag: flag if the payment was made in the interim phase

“Forma_de_pago”	Payment method: methodology or channel through which the payment was made

Source: Own elaboration

b. File extract example

Figure 104. File “Cobros” extract example

COB_NUMERO	IUC	SERIE	NUMERO	Cartera	Cli_negocio	Collection_date	IMPORTE	INTERIM	Forma_de_pago
17395503	22503958	1513	190	G	I	5/4/2022	100		Friendly
17754962	22503958	1513	190	G	I	2/6/2022	100		Friendly
17593191	22503958	1513	190	G	I	10/5/2022	100		Friendly
17977115	22503958	1513	190	G	I	5/7/2022	100		Friendly
18394329	22503958	1513	190	G	I	6/9/2022	100		Friendly
18141004	22503958	1513	190	G	I	2/8/2022	100		Friendly
18598903	22503958	1513	190	G	I	7/10/2022	100		Friendly
18827147	22503958	1513	190	G	I	14/11/2022	100		Friendly
17851268	22504717	1513	949	G	I	22/6/2022	2001,07		Friendly
17851269	22504717	1513	949	G	I	22/6/2022	-1363		Friendly

Source: Own elaboration

14- **File “ICC – indicador de confianza del consumidor” – Consumer Confidence Index**: file with the information about the consumer confidence index indicators, downloaded from the INE website.

a. File fields

Table 112. Fields in the File “ICC – indicador de confianza del consumidor”

Field	Description
“Fecha”	Date: month and year of the indicator
“ICC”	Consumer Confidence Index
“Ind_Sit_Actual”	Consumer Confidence Index actual situation
“Ind_Expectativas”	Consumer Confidence Index expectations

Source: Own elaboration

b. File extract example

Figure 105. File "ICC – indicador de confianza del consumidor" extract example

Fecha	ICC	Ind_Sit_Actual	Ind_Expectativas
01/01/2014	77,7	53,8	93,4
01/02/2014	71,5	62	88
01/03/2014	76,3	55	92,2
01/04/2014	82	60,4	97,7
01/05/2014	84,9	66,2	97,6
01/06/2014	89,3	72,2	104,2
01/07/2014	88,9	74,4	100,8
01/08/2014	87,7	77	98,4
01/09/2014	89,3	77	102,2
01/10/2014	86,8	76,3	99,3

Source: Own elaboration

15- File "IPC_IndGeneral_VarAnual" – General Consumer Price Index on Annual Variation: file with the information about the General Consumer Price Index indicators on annual variation, downloaded from the INE website.

a. File fields

Table 113. Fields in the File "IPC_IndGeneral_VarAnual"

Field	Description
"Fecha"	Date: month and year of the indicator
"IPC_IndGeneral_VarAnual"	General Consumer Price Index on annual variation

Source: Own elaboration

b. File extract example

Figure 106. File "IPC_IndGeneral_VarAnual" extract example

Fecha	IPC_IndGeneral_VarAnual
01/07/2023	2,3
01/06/2023	1,9
01/05/2023	3,2
01/04/2023	4,1
01/03/2023	3,3
01/02/2023	6
01/01/2023	5,9
01/12/2022	5,7
01/11/2022	6,8
01/10/2022	7,3

Source: Own elaboration

16- **File "IPC_IndGeneral_VarMensual" – General Consumer Price Index on Monthly Variation**: file with the information about the General Consumer Price Index indicators on monthly variation, downloaded from the INE website.

a. File fields

Table 114. Fields in the File "IPC_IndGeneral_VarMensual"

Field	Description
"Fecha"	Date: month and year of the indicator
"IPC_IndGeneral_VarMensual"	General Consumer Price Index on monthly variation

Source: Own elaboration

b. File extract example

Figure 107. File "IPC_IndGeneral_VarMensual" extract example

Fecha	IPC_IndGeneral_VarMensual
01/07/2023	0,1
01/06/2023	0,6
01/05/2023	0
01/04/2023	0,6
01/03/2023	0,4
01/02/2023	0,9
01/01/2023	-0,2
01/12/2022	0,2
01/11/2022	-0,1
01/10/2022	0,3

Source: Own elaboration

17- **File "IPC_IndSubyacente_VarAnual" – Underlying Consumer Price Index on Annual Variation**: file with the information about the Underlying Consumer Price Index indicators on annual variation, downloaded from the INE website.

a. File fields

Table 115. Fields in the File "IPC_IndSubyacente_VarAnual"

Field	Description
"Fecha"	Date: month and year of the indicator
"IPC_IndSubyacente_VarAnual"	Underlying Consumer Price Index on annual variation

Source: Own elaboration

b. File extract example

Figure 108. File "IPC_IndSubyacente_VarAnual" extract example

Fecha	IPC_IndSubyacente_VarAnual
01/07/2023	6,2
01/06/2023	5,9
01/05/2023	6,1
01/04/2023	6,6
01/03/2023	7,5
01/02/2023	7,6
01/01/2023	7,5
01/12/2022	7
01/11/2022	6,3
01/10/2022	6,2

Source: Own elaboration

18- **File "IPC_IndSubyacente_VarMensual" – Underlying Consumer Price Index on Monthly Variation**: file with the information about the Underlying Consumer Price Index indicators on monthly variation, downloaded from the INE website.

a. File fields

Table 116. Fields in the File "IPC_IndSubyacente_VarMensual"

Field	Description
"Fecha"	Date: month and year of the indicator
"IPC_IndSubyacente_VarMensual"	Underlying Consumer Price Index on monthly variation

Source: Own elaboration

b. File extract example

Figure 109. File "IPC_IndSubyacente_VarMensual" extract example

Fecha	IPC_IndSubyacente_VarMensual
01/07/2023	0
01/06/2023	0,5
01/05/2023	0,2
01/04/2023	1
01/03/2023	0,7
01/02/2023	0,7
01/01/2023	-0,2
01/12/2022	0,9
01/11/2022	0,7
01/10/2022	1

Source: Own elaboration

19- **File "Paro_TasaActividad_VarAnual" – Activity Rate on annual variation**: file with the information about the Activity Rate indicators on annual variation, downloaded from the INE website.

a. File fields

Table 117. Fields in the File "Paro_TasaActividad_VarAnual"

Field	Description
"Periodo"	Period: quarter and year of the indicator
"Paro_TasaActividad_VarAnual"	Activity Rate on annual variation

Source: Own elaboration

b. File extract example

Figure 110. File "Paro_TasaActividad_VarAnual" extract example

PERIODO	Paro_TasaActividad_VarAnual
2023T2	0,26
2023T1	0,05
2022T4	-0,13
2022T3	-0,28
2022T2	0,14
2022T1	0,81
2021T4	0,46
2021T3	1,3
2021T2	3,03
2021T1	-0,49

Source: Own elaboration

20- **File "Paro_TasaParo_VarAnual" – Unemployment Rate on annual variation**: file with the information about the Unemployment Rate indicators on annual variation, downloaded from the INE website.

a. File fields

Table 118. Fields in the File "Paro_TasaParo_VarAnual"

Field	Description
"Periodo"	Period: quarter and year of the indicator
"Paro_TasaParo_VarAnual"	Unemployment Rate on annual variation

Source: Own elaboration

b. File extract example

Figure 111. File "Paro_TasaParo_VarAnual" extract example

PERIODO	Paro_TasaParo_VarAnual
2023T2	-0,89
2023T1	-0,38
2022T4	-0,45
2022T3	-1,9
2022T2	-2,78
2022T1	-2,33
2021T4	-2,8
2021T3	-1,69
2021T2	-0,06
2021T1	1,58

Source: Own elaboration

21- **File "PIB_EmpleoEquivTiempoCompleto_VarAnual" – GDP Full-Time Equivalent Employment on annual variation**: file with the information about the GDP Full-Time Equivalent Employment indicators on annual variation, downloaded from the INE website.

a. File fields

Table 119. Fields in the File "PIB_EmpleoEquivTiempoCompleto_VarAnual"

Field	Description
"Periodo"	Period: quarter and year of the indicator
"PIB_EmpleoEquivTiempoCompleto_VarAnual"	GDP Full-Time Equivalent Employment on annual variation

Source: Own elaboration

b. File extract example

Figure 112. File "PIB_EmpleoEquivTiempoCompleto_VarAnual" extract example

PERIODO	PIB_EmpleoEquivTiempoCompleto_VarAnual
2023T2	2,877
2023T1	2,287
2022T4	2,077
2022T3	2,826
2022T2	5,009
2022T1	5,322
2021T4	5,999
2021T3	6,364
2021T2	18,906
2021T1	-2,748

Source: Own elaboration

22- **File "PIB_EmpleoEquivTiempoCompleto_VarTrim" – GDP Full-Time Equivalent Employment on quarterly variation**: file with the information about the GDP Full-Time

Equivalent Employment indicators on quarterly variation, downloaded from the INE website.

a. File fields

Table 120. Fields in the File "PIB_EmpleoEquivTiempoCompleto_VarTrim"

Field	Description
"Periodo"	Period: quarter and year of the indicator
"PIB_EmpleoEquivTiempoCompleto_VarTrim"	GDP Full-Time Equivalent Employment on quarterly variation

Source: Own elaboration

b. File extract example

Figure 113. File "PIB_EmpleoEquivTiempoCompleto_VarTrim" extract example

PERIODO	PIB_EmpleoEquivTiempoCompleto_VarTrim
2023T2	1,306
2023T1	0,21
2022T4	0,177
2022T3	1,16
2022T2	0,725
2022T1	0,004
2021T4	0,911
2021T3	3,308
2021T2	1,025
2021T1	0,647

Source: Own elaboration

23- **File "PIB_PreciosMercado_VarAnual" – GDP Market Prices on annual variation**: file with the information about the GDP Market Prices indicators on annual variation, downloaded from the INE website.

a. File fields

Table 121. Fields in the File "PIB_PreciosMercado_VarAnual"

Field	Description
"Periodo"	Period: quarter and year of the indicator
"PIB_PreciosMercado_VarAnual"	GDP Market Prices on annual variation

Source: Own elaboration

b. File extract example

Figure 114. File "PIB_PreciosMercado_VarAnual" extract example

PERIODO	PIB_PreciosMercado_VarAnual
2023T2	1,753
2023T1	4,156
2022T4	3,022
2022T3	4,922
2022T2	7,766
2022T1	6,256
2021T4	6,625
2021T3	4,202
2021T2	17,875
2021T1	-4,43

Source: Own elaboration

24- **File "PIB_PreciosMercado_VarTrim" – GDP Market Prices on quarterly variation**: file with the information about the GDP Market Prices indicators on quarterly variation, downloaded from the INE website.

a. File fields

Table 122. Fields in the File "PIB_PreciosMercado_VarTrim"

Field	Description
"Periodo"	Period: quarter and year of the indicator
"PIB_PreciosMercado_VarTrim"	GDP Market Prices on quarterly variation

Source: Own elaboration

b. File extract example

Figure 115. File "PIB_PreciosMercado_VarTrim" extract example

PERIODO	PIB_PreciosMercado_VarTrim
2023T2	0,42
2023T1	0,52
2022T4	0,414
2022T3	0,387
2022T2	2,792
2022T1	-0,574
2021T4	2,265
2021T3	3,108
2021T2	1,352
2021T1	-0,229

Source: Own elaboration

APPENDIX 2.- DATA CLEANING AND NORMALIZATION DETAIL

2.1.- Creditor's grouping detail

In the following table is shown the groupings and transformations of the creditor's names.

Table 123. Creditor's names grouping

Original creditor name	Grouped creditor name	Accounts
ORANGE	ORANGE	425.697
JAZZTEL	JAZZTEL	236.757
	DESC	105.249
CAIXABANK	CAIXABANK	94.553
BANKIA / CAJA MADRID	BANKIA / CAJA MADRID	68.298
SANTANDER	SANTANDER	55.459
BANCO SANTANDER	SANTANDER	44.434
ING BANK N.V., SUCURSAL EN ESPAA	ING	38.527
CAJA DUERO	CAJA/RURAL	33.890
BANKIA	BANKIA	30.003
UNICAJA	UNICAJA	28.755
CATALUNYA CAIXA	CATALUNYA CAIXA	28.392
CAJASOL	CAJA/RURAL	25.423
NULL	DESC	24.626
LIBERBANK	LIBERBANK	19.016
………		
5921	DESC	1
9504	DESC	1
866	DESC	1
C.CAMPO C.RURAL SCC VALENCIA	CAJA/RURAL	1
4627	DESC	1

Source: Own elaboration

2.2.- Product type grouping detail

In the following table is shown the groupings and transformations of the product types.

Table 124. Product type grouping

Original product type	Grouped product type	Accounts
TELEFONIA	06_TELCO	661.817
PRESTAMO CONSUMO	02_PRESTAMO	145.140
TARJETA DE CREDITO	01_TARJETA CREDITO	109.829
DESCUBIERTO CTA/CTE	04_DESCUBIERTO	86.793
CUENTA CORRIENTE	03_CUENTA CORRIENTE	75.551
OVERDRAFTS	04_DESCUBIERTO	44.352
POLIZA DE PRESTAMO PERS	02_PRESTAMO	42.350

DESCUBIERTO	04_DESCUBIERTO	38.055
TARJETA DE CREDITO	01_TARJETA CREDITO	35.792
PRESTAMO	02_PRESTAMO	32.307
LOANS	02_PRESTAMO	27.002
TARJETA DE CREDITO / CAJA MADRID / BANKI	01_TARJETA CREDITO	19.378
CUENTA DE CREDITO	03_CUENTA CORRIENTE	18.703
AVALES	08_AVAL	16.109
	15_OTROS	14.268
………		
/ / ELEC	15_OTROS	1
CREDITO	10_CREDITO	1
BFA/CREDIT/23000/	10_CREDITO	1
BANK GUARANTEE	12_GARANTIA	1
GARANTIAS	12_GARANTIA	1

Source: Own elaboration

APPENDIX 3.- EXPLANATORY VARIABLES DETAIL

3.1.- Variable "IN01_Cartera_G" – Portfolio

Description

Name of the original portfolio. A synthetic variable was built to group the different portfolios. Due to business needs of the Investment Fund, at the beginning of the management the same portfolio was split with similar but different names. So, for modelling purposes, we regrouped them in the original unique portfolios.

This variable was originally included in the data set, as normally shows correlation with the bigger or lower probability to collect (due to different characteristics of debtor, debt, collections activities of the original creditor, credit policies…, between the different portfolios). But in our modelling, finally this variable was not relevant. This is due to having included in the data set all the variables intrinsically linked to the portfolio, as type of product, debt age, amounts unpaid…

Individuals' details

30 different portfolios are included. The two biggest portfolios sum up for 27% of the total accounts and the seven biggest portfolios compromises 57% of the total accounts.

SMEs details

23 different portfolios are included in this data ser. The two biggest portfolios sum up for more than 34% of the total accounts and the nine biggest portfolios compromises 74% of the total accounts.

Discrimination bivariant analysis for Individuals and SMEs

We observe that there is correlation between the portfolio and the payment in the following 30 days (the distribution of accounts and payments is not similar for the different portfolios). Certain portfolios have a higher probability for a payment than others, both for Individuals and SMEs. The information value (IV) of this variable is 0.47 for Individuals and 0.46 for SMEs.

3.2.- Variable "IN02_Vertical" – Vertical segment

Description

Vertical segment of each portfolio. As the data sets have been split in Individuals and SMEs for the model's construction, in each data set the value for all the accounts is the same: in the Individuals data set the vertical is Individuals and in the SMEs data set is SMEs. For this same reason, this variable is not relevant for the modelling exercise. If a unique data set would have been built, this variable would be relevant for the modelling.

Individuals' and SMEs details

Table 125. Distribution of Individuals accounts by vertical

IN02_Vertical	Record_Count
Individual	676.028

Source: Own elaboration

Table 126. Distribution of SMEs accounts by vertical

IN02_Vertical	Record_Count
SME	195.962

Source: Own elaboration

3.3.- Variable "IN03_Originador" – Creditor

Description

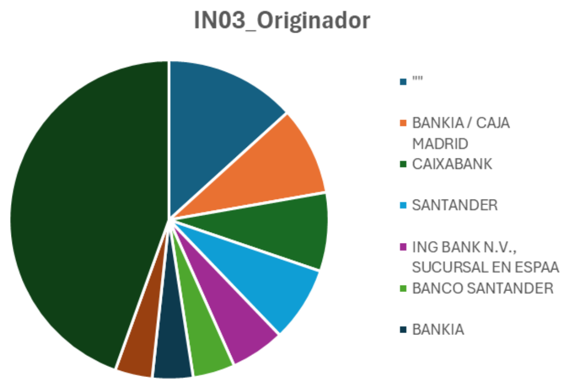
Name of the original creditor of the debt (without grouping and before the cleaning of null cases). This variable was included in the data set, as normally shows correlation with the bigger or lower probability to collect (due to different characteristics of debtor, debt, collections activities, credit policies…, between the different creditors). Finally, this variable by itself was not relevant for the modelling, but the grouping of this category (variable "IN11_Originador_G" – Creditor grouped) was relevant for the modelling in all the models built.

Individuals' details

More than one thousand (1.093) different creditors were originally in the data set. The biggest creditor represents 9% of the total accounts and the five biggest creditors compromises 38% of the total accounts (being two of the six biggest Banks in the Spanish market). 13% of the accounts have null values on the creditor field.

Table 127 and Figure 116. Distribution of Individuals accounts by creditor

IN03_Originador	Record_Count
	89.731
BANKIA / CAJA MADRID	60.580
CAIXABANK	53.916
SANTANDER	51.506
ING BANK N.V., SUCURSAL EN ESPAA	37.119
BANCO SANTANDER	28.765
BANKIA	27.876
CATALUNYA CAIXA	25.565
CAJA DUERO	24.286



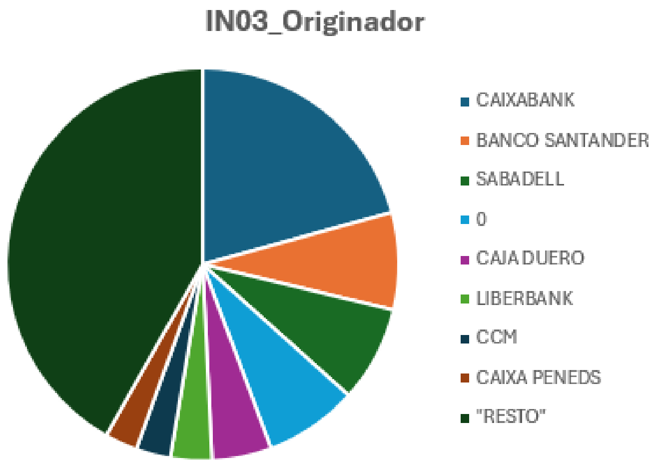
Source: Own elaboration

SMEs details

More than one thousand (1,093) different creditors were originally in the data set. The biggest creditor represents 21% of the total accounts and the three biggest creditors compromises 37% of the total accounts (being three of the six biggest Banks in the Spanish market). 7% of the accounts have null values on the creditor field.

Table 128 and Figure 117. Distribution of SMEs accounts by creditor

IN03_Originador	Record_Count
CAIXABANK	40.635
BANCO SANTANDER	15.645
SABADELL	15.493
	15.085
CAJA DUERO	9.604
LIBERBANK	6.712
CCM	5.567
CAIXA PENEDS	5.281
UNICAJA	5.009



Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

We observe that there is correlation between the creditor and the payment in the following 30 days (the distribution of accounts and payments is not similar for the different creditors). Certain creditors have a higher probability for a payment than others, both for Individuals and SMEs. In the grouped version of this variable (variable "IN11_Originador_G" – Creditor grouped), we detail which are these creditors with higher probability. The information value (IV) of this variable is 0.35 for Individuals and 0.31 for SMEs.

3.4.- Variable "IN06_Saldo_INI_INT_G" – Initial interest debt amount

Description

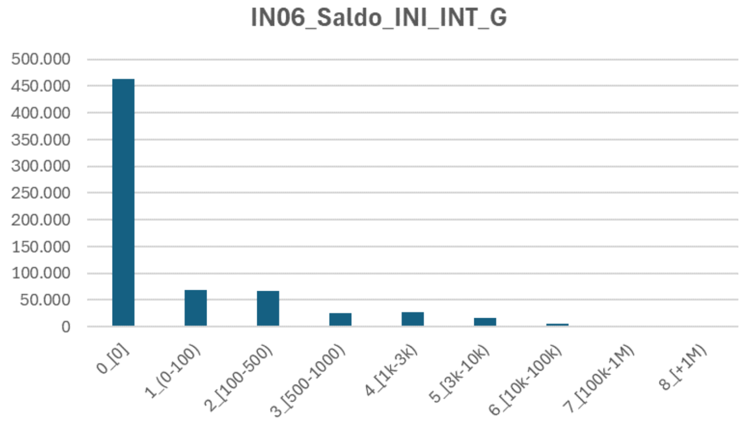
Initial interest outstanding debt amount of the account, grouped in ranges or buckets. This variable was included in the data set, as normally shows correlation with the probability to collect. Lower amount shows higher probability to collect, and higher amounts show lower probability. Finally, this variable was relevant in all the models built, except in the Individuals Post Covid-19 one.

Individuals' details

The distribution of accounts is highly concentrated on the first bucket of 0€ initial interest debt amount (69% of total accounts). The rest of account distributes in a decreasing trend according to the increase of the interest's amount.

Table 129 and Figure 118. Distribution of Individuals accounts by initial interest debt amount

IN06_Saldo_INI_INT_G	Record_Count
0_[0]	464.163
1_(0-100)	69.647
2_[100-500)	66.540
3_[500-1000)	26.203
4_[1k-3k)	26.566
5_[3k-10k)	16.574
6_[10k-100k)	6.142
7_[100k-1M)	188
8_[+1M)	5



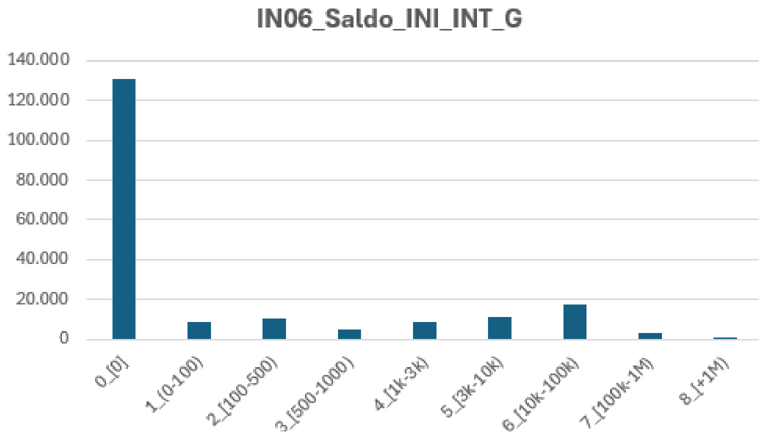
Source: Own elaboration

SMEs details

The distribution of accounts is highly concentrated on the first bucket of 0€ initial interest debt amount (67% of total accounts). The rest of account distributes homogeneously in the rest of buckets.

Table 130 and Figure 119. Distribution of SMEs accounts by initial interest debt amount

IN06_Saldo_INI_INT_G	Record_Count
0_[0]	130.692
1_(0-100)	8.692
2_[100-500)	10.378
3_[500-1000)	4.778
4_[1k-3k)	8.912
5_[3k-10k)	11.334
6_[10k-100k)	17.442
7_[100k-1M)	3.575
8_[+1M)	159



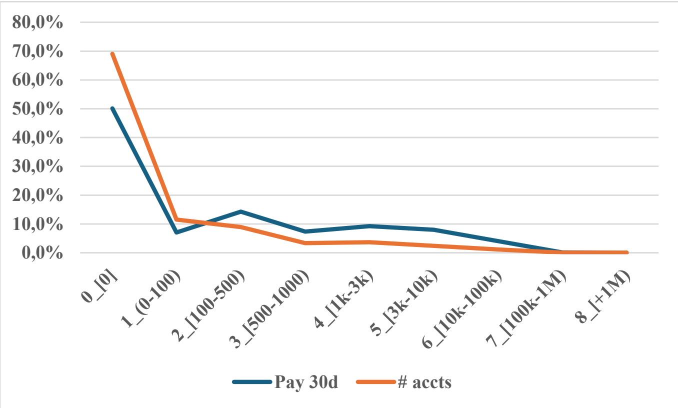
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

In this case, we observe that there is certain correlation between the initial interest debt amount and the payment in the following 30 days (the distribution of accounts and payments is not similar for the different ranges of the initial interest debt amount). In Individuals, the probability for a payment is higher in all the ranges above 100. And in SMEs, it is a similar situation, with higher probabilities in the ranges 500-1000, 1-3k, 3-10k, 10-100k and 100k-1M. The information value (IV) of this variable is 0.30 for Individuals and 0.27 for SMEs.

Table 131 and Figure 120. Bivariant analysis initial interest debt amount – payment 30 days for Individuals

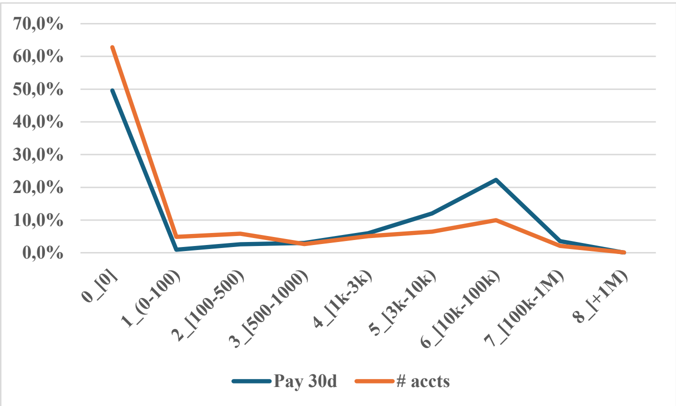
Initial interest debt amount	Pay 30d	# accts
0_[0]	50,1%	69,1%
1_(0-100)	7,0%	11,5%
2_[100-500)	14,3%	8,9%
3_[500-1000)	7,3%	3,4%
4_[1k-3k)	9,2%	3,6%
5_[3k-10k)	7,9%	2,4%
6_[10k-100k)	4,1%	1,1%
7_[100k-1M)	0,1%	0,0%
8_[+1M)	0,0%	0,0%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 132 and Figure 121. Bivariant analysis initial interest debt amount – payment 30 days for SMEs

Initial interest debt amount	Pay 30d	# accts
0_[0]	49,6%	62,8%
1_(0-100)	0,9%	4,9%
2_[100-500)	2,6%	5,8%
3_[500-1000)	3,0%	2,7%
4_[1k-3k)	6,0%	5,1%
5_[3k-10k)	12,0%	6,5%
6_[10k-100k)	22,3%	10,0%
7_[100k-1M)	3,6%	2,1%
8_[+1M)	0,0%	0,1%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.5.- Variable "IN08_Saldo_INI_OTH_G" – Initial other debt amount

Description

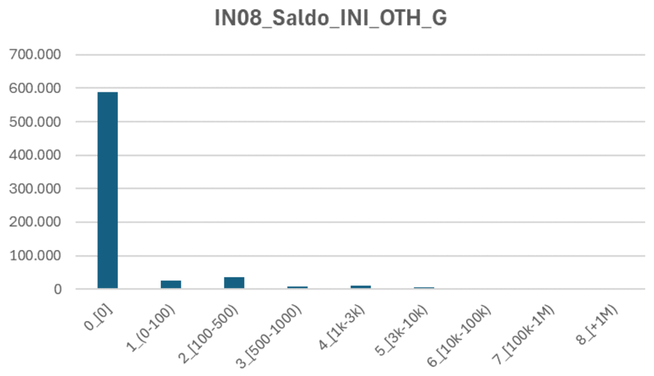
Initial other outstanding debt amount of the account (not interest nor principal amounts), grouped in ranges or buckets. This variable was included in the data set, as normally shows correlation with the probability to collect. Lower amount shows higher probability to collect, and higher amounts show lower probability. Finally, this variable was relevant in our modelling for the Individuals Global and the Individuals Pre Covid-19 models, but it was not relevant for the SMEs models.

Individuals' details

The distribution of accounts is highly concentrated on the first bucket of 0€ initial other debt amount (87% of total accounts). The rest of account distributes homogeneously in the rest of buckets.

Table 133 and Figure 122. Distribution of Individuals accounts by initial other debt amount

IN08_Saldo_INI_OTH_G	Record_Count
0_[0]	587.775
1_(0-100)	25.153
2_[100-500)	36.400
3_[500-1000)	8.338
4_[1k-3k)	11.593
5_[3k-10k)	6.013
6_[10k-100k)	752
7_[100k-1M)	4
8_[+1M)	0



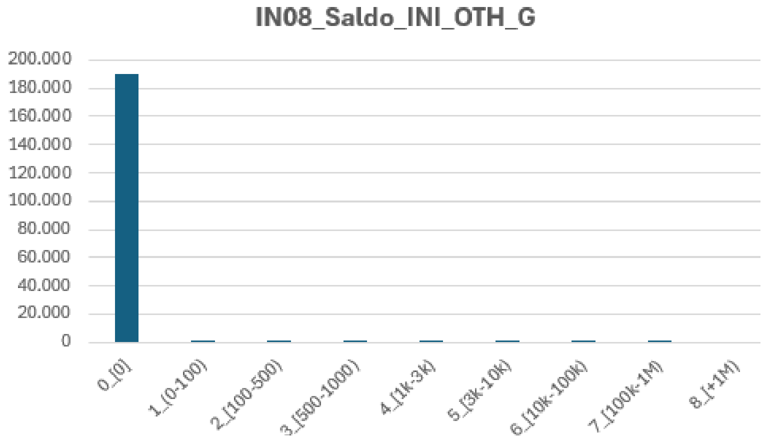
Source: Own elaboration

SMEs details

The distribution of accounts is highly concentrated on the first bucket of 0€ initial other debt amount (97% of total accounts). The rest of account distributes homogeneously in the rest of buckets.

Table 134 and Figure 123. Distribution of SMEs accounts by initial other debt amount

IN08_Saldo_INI_OTH_G	Record_Count
0_[0]	189.875
1_(0-100)	984
2_[100-500)	1.623
3_[500-1000)	669
4_[1k-3k)	1.479
5_[3k-10k)	887
6_[10k-100k)	430
7_[100k-1M)	15
8_[+1M)	0



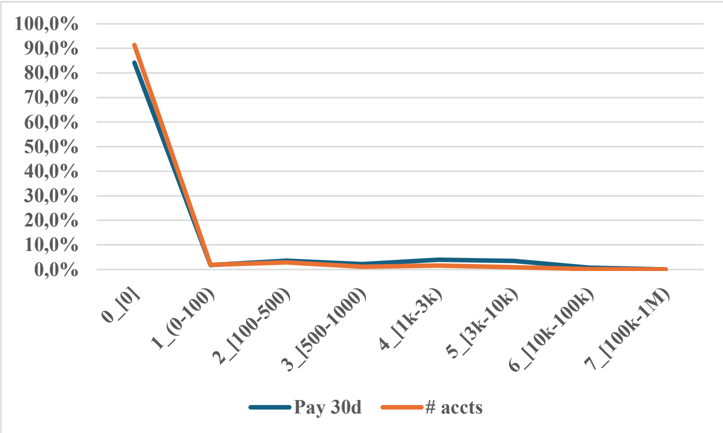
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

In this case, we observe that there is certain correlation between the initial other debt amount and the payment in the following 30 days (the distribution of accounts and payments is not similar for the different ranges of the initial other debt amount). In both Individuals and SMEs, the probability for a payment is higher in all the ranges above 100; but in Individuals especially higher for the 1-3k, 3-10k and 10-100k ranges, and in SMEs specially in the ranges 500-1000, 1-3k and 3-10k. The information value (IV) of this variable is 0.08 for Individuals and 0.05 for SMEs.

Table 135 and Figure 124. Bivariant analysis initial other debt amount – payment 30 days for Individuals

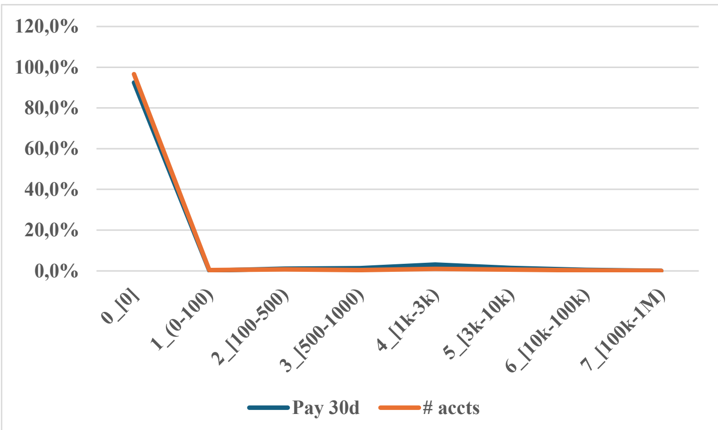
Initial other debt amount	Pay 30d	# accts
0_[0]	84,2%	91,4%
1_(0-100)	1,8%	1,9%
2_[100-500)	3,5%	2,9%
3_[500-1000)	2,2%	1,1%
4_[1k-3k)	4,0%	1,6%
5_[3k-10k)	3,5%	0,9%
6_[10k-100k)	0,7%	0,1%
7_[100k-1M)	0,0%	0,0%
Total general	100,0%	100,0%



Source: Own elaboration

Table 136 and Figure 125. Bivariant analysis initial other debt amount – payment 30 days for SMEs

Initial other debt amount	Pay 30d	# accts
0_[0]	92,5%	96,6%
1_(0-100)	0,1%	0,4%
2_[100-500)	1,0%	0,7%
3_[500-1000)	1,3%	0,4%
4_[1k-3k)	3,1%	1,0%
5_[3k-10k)	1,4%	0,6%
6_[10k-100k)	0,5%	0,3%
7_[100k-1M)	0,1%	0,0%
Total general	100,0%	100,0%



Source: Own elaboration

3.6.- Variable "IN09_Mes_Impago" – Month and year of the original debt

Description

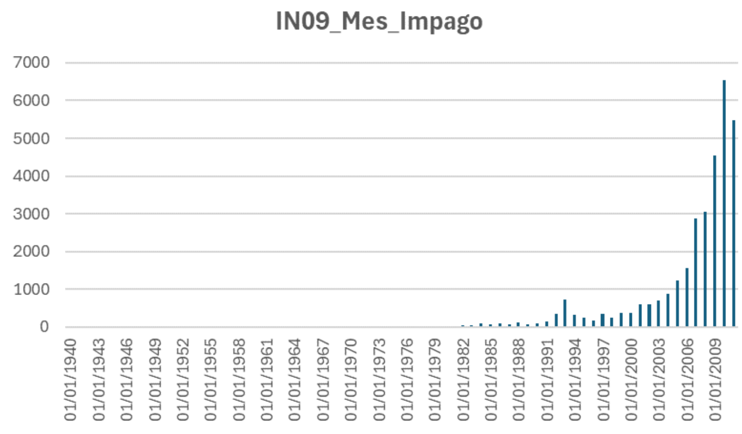
Month and year when the debt was originated. This variable was included in the data set, as normally shows correlation with the probability to collect. Older debt shows lower probability to collect, and younger debt shows higher probability. Despite this typical behaviour, this variable was finally not relevant for the models built.

Individuals' and SMEs details

The distribution of accounts is highly concentrated after 2007-2008 period (financial crisis linked to the mortgages "bubble"), and there is an important peak in 2020 (linked to the Covid-19 crisis). Records before 2007 are residual.

Table 137 and Figure 126. Distribution of Individuals accounts by date of the original debt

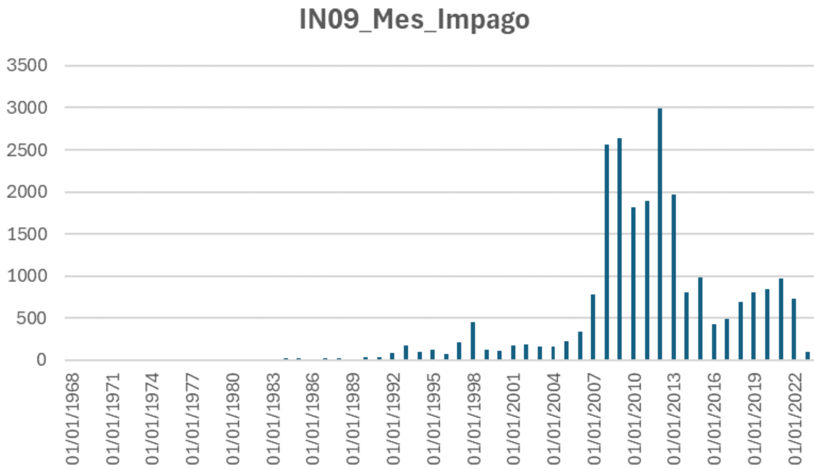
IN09_Mes_Impago	Record_Count
01/04/2012	2.995
01/01/2009	2.634
01/12/2008	2.561
01/01/2012	2.513
01/03/2012	2.314
01/10/2008	2.293
01/03/2009	2.264
01/07/2012	2.224
01/04/2009	2.221
01/02/2009	2.196
01/11/2008	2.156
01/06/2012	2.148
01/05/2012	2.073



Source: Own elaboration

Table 138 and Figure 127. Distribution of SMEs accounts by date of the original debt

IN09_Mes_Impago	Record_Count
01/04/2012	2.995
01/01/2009	2.634
01/12/2008	2.561
01/01/2012	2.513
01/03/2012	2.314
01/10/2008	2.293
01/03/2009	2.264
01/07/2012	2.224
01/04/2009	2.221
01/02/2009	2.196
01/11/2008	2.156
01/06/2012	2.148
01/05/2012	2.073

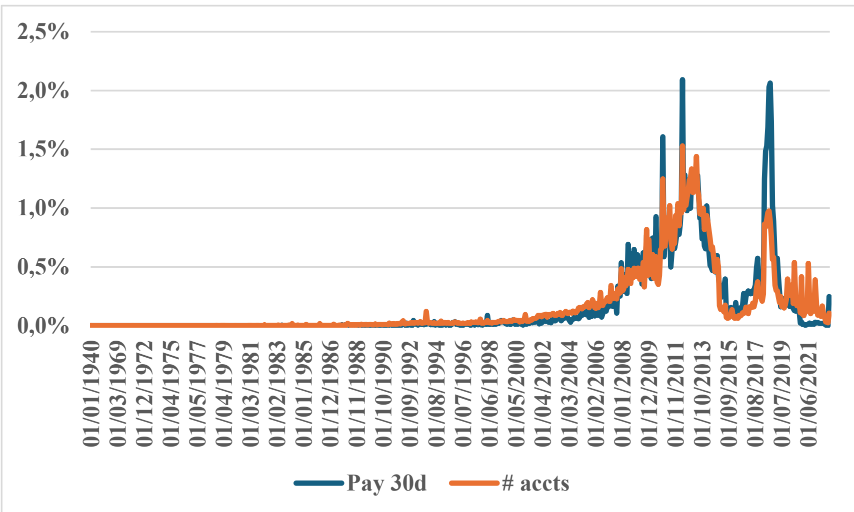


Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

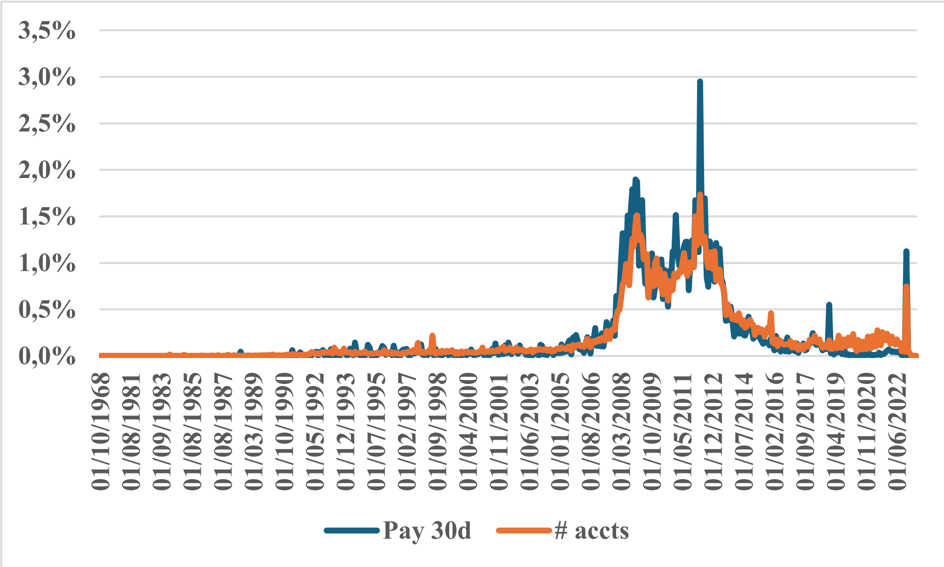
For this variable, we observe that there is certain correlation between the month and year of the original debt and the payment in the following 30 days (the distribution of accounts and payments is not similar for the different months and years of the original debt). In both Individuals and SMEs, the probability for a payment varies according to the month and year of the original debt. The information value (IV) of this variable is 0.25 for Individuals and 0.25 for SMEs.

Figure 128. Bivariant analysis month and year of the original debt – payment 30 days for Individuals



Source: Own elaboration

Figure 129. Bivariant analysis month and year of the original debt – payment 30 days for SMEs



Source: Own elaboration

3.7.- Variable "IN10_Anio_Impago" – Year of the original debt

Description

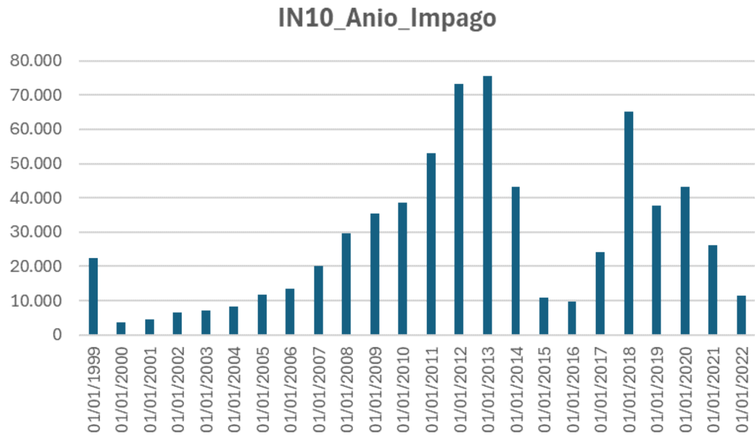
Year when the debt was originated. This variable was included in the data set, as normally shows correlation with the probability to collect. Older debt shows lower probability to collect, and younger debt shows higher probability. Despite this typical behaviour, this variable was finally not relevant for the models built.

Individuals' and SMEs details

The distribution of accounts is highly concentrated after 2007-2008 period (financial crisis linked to the mortgages "bubble"), and there is an important peak in 2020 (linked to the Covid-19 crisis). Values before 2000 were grouped as they have a similar behaviour and in order to have more materiality in this bucket.

Table 139 and Figure 130. Distribution of Individuals accounts by year of the original debt

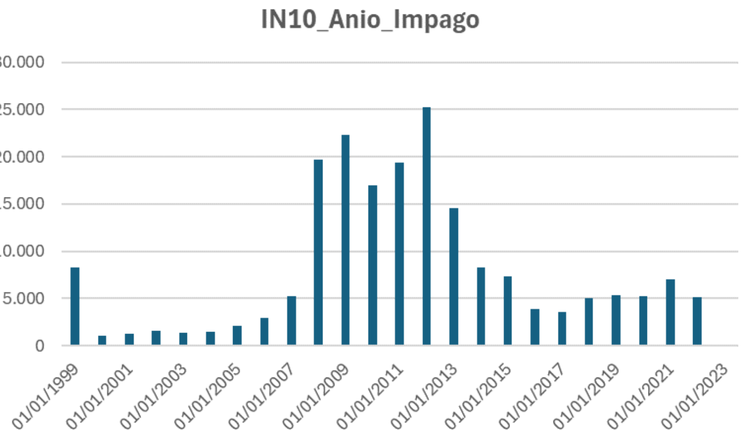
IN10_Anio_Impago	Record_Count
01/01/2013	75.515
01/01/2012	73.197
01/01/2018	65.240
01/01/2011	53.156
01/01/2014	43.316
01/01/2020	43.264
01/01/2010	38.722
01/01/2019	37.605
01/01/2009	35.498



Source: Own elaboration

Table 140 and Figure 131. Distribution of SMEs accounts by year of the original debt

IN10_Anio_Impago	Record_Count
01/01/2012	25.247
01/01/2009	22.307
01/01/2008	19.727
01/01/2011	19.393
01/01/2010	16.990
01/01/2013	14.592
01/01/2014	8.303
01/01/1999	8.286
01/01/2015	7.368



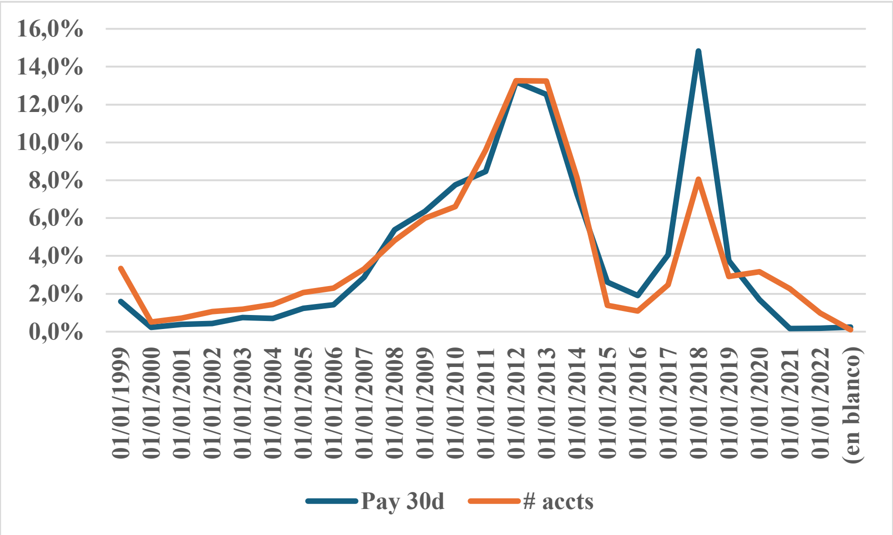
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

In this case, we observe that there is certain correlation between the year of the original debt and the payment in the following 30 days (the distribution of accounts and payments is not similar for the different years of the original debt). In general, the older the year of the original debt, the lower the for a payment. In Individuals, the higher probability of payments is for the years in the 2015-2019 range. In the case of SMEs, this probability is higher for the 2007-2013 years. The information value (IV) of this variable is 0.19 for Individuals and 0.12 for SMEs.

Table 141 and Figure 132. Bivariant analysis year of the original debt – payment 30 days for Individuals

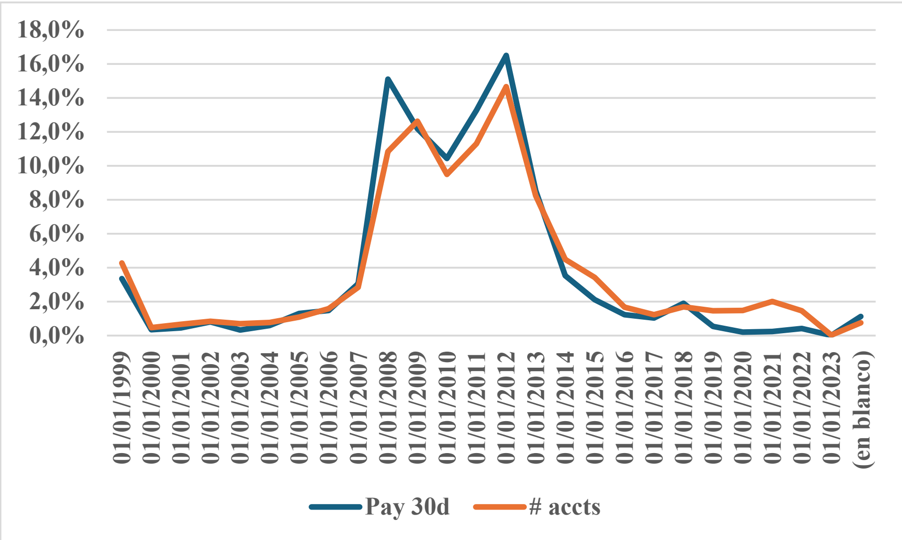
Year of the original debt	Pay 30d	# accts
01/01/1999	1,6%	3,3%
01/01/2000	0,2%	0,5%
01/01/2001	0,4%	0,7%
01/01/2002	0,4%	1,1%
01/01/2003	0,7%	1,2%
01/01/2004	0,7%	1,4%
01/01/2005	1,2%	2,1%
01/01/2006	1,4%	2,3%
01/01/2007	2,9%	3,3%
01/01/2008	5,4%	4,8%
01/01/2009	6,4%	6,0%
01/01/2010	7,8%	6,6%
01/01/2011	8,5%	9,6%
01/01/2012	13,2%	13,3%
01/01/2013	12,5%	13,2%
01/01/2014	7,3%	8,1%
01/01/2015	2,6%	1,4%
01/01/2016	1,9%	1,1%
01/01/2017	4,1%	2,5%
01/01/2018	14,8%	8,1%
01/01/2019	3,7%	2,9%
01/01/2020	1,7%	3,2%
01/01/2021	0,2%	2,3%
01/01/2022	0,2%	1,0%
(en blanco)	0,2%	0,1%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 142 and Figure 133. Bivariant analysis year of the original debt – payment 30 days for SMEs

Year of the original debt	Pay 30d	# accts
01/01/1999	3,4%	4,3%
01/01/2000	0,3%	0,5%
01/01/2001	0,4%	0,7%
01/01/2002	0,8%	0,8%
01/01/2003	0,3%	0,7%
01/01/2004	0,6%	0,8%
01/01/2005	1,3%	1,1%
01/01/2006	1,5%	1,6%
01/01/2007	3,1%	2,8%
01/01/2008	15,1%	10,8%
01/01/2009	12,2%	12,6%
01/01/2010	10,4%	9,5%
01/01/2011	13,3%	11,3%
01/01/2012	16,5%	14,7%
01/01/2013	8,5%	8,2%
01/01/2014	3,5%	4,5%
01/01/2015	2,1%	3,4%
01/01/2016	1,2%	1,7%
01/01/2017	1,0%	1,2%
01/01/2018	1,9%	1,7%
01/01/2019	0,5%	1,5%
01/01/2020	0,2%	1,5%
01/01/2021	0,2%	2,0%
01/01/2022	0,4%	1,5%
01/01/2023	0,0%	0,0%
(en blanco)	1,1%	0,7%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.8.- Variable "IN11_Originador_G" – Creditor grouped

Description

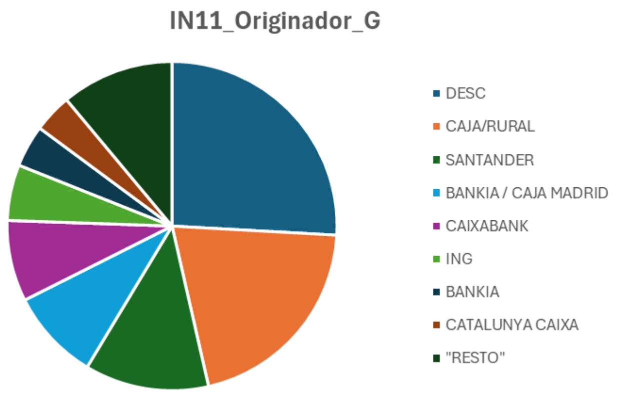
Name of the original creditor of the debt, grouped and after cleaning of null cases. This variable was included in the data set, as normally shows correlation with the bigger or lower probability to collect (due to different characteristics of debtor, debt, collections activities, credit policies…, between the different creditors). And in our modelling this variable was relevant for all the models built in Individuals and SMEs.

Individuals' details

26 different groups of creditors were defined in the data set. 39% of the total accounts belongs to some of the main banks in the Spanish market, 21% of the accounts have its origin in saving banks or rural saving banks, and 26% have unknown values.

Table 143 and Figure 134. Distribution of Individuals accounts by creditor grouped

IN11_Originador_G	Record_Count
DESC	174.958
CAJA/RURAL	138.987
SANTANDER	82.269
BANKIA / CAJA MADRID	60.580
CAIXABANK	53.916
ING	37.119
BANKIA	27.876
CATALUNYA CAIXA	25.565
UNICAJA	23.722



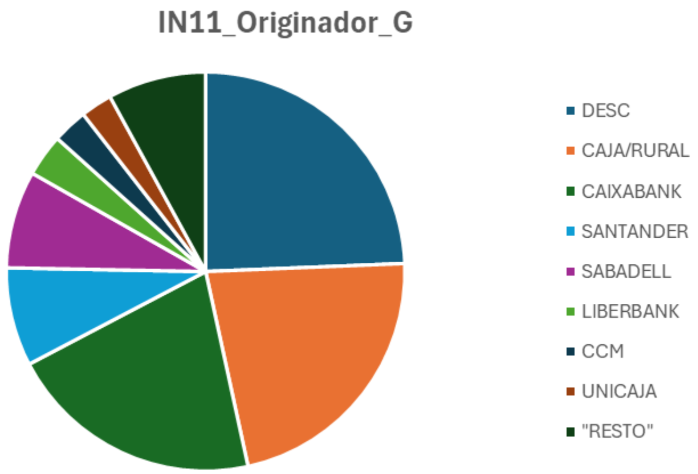
Source: Own elaboration

SMEs details

26 different groups of creditors were defined in the data set. 35% of the total accounts belongs to some of the main banks in the Spanish market, 25% of the accounts have its origin in saving banks or rural saving banks, and 25% have unknown values.

Table 144 and Figure 135. Distribution of SMEs accounts by creditor grouped

IN11_Originador_G	Record_Count
DESC	47.738
CAJA/RURAL	43.540
CAIXABANK	40.635
SANTANDER	15.645
SABADELL	15.493
LIBERBANK	6.712
CCM	5.567
UNICAJA	5.009
BANKIA / CAJA MADRID	4.824
BANCO DE VALENCIA	3.006
BANCAJA	2.443
SA NOSTRA	1.680
ING	1.408
BMN	1.137
BANCO POPULAR	416
BARCLAYS BANK	356
BANKIA	257
BANKPIME	96

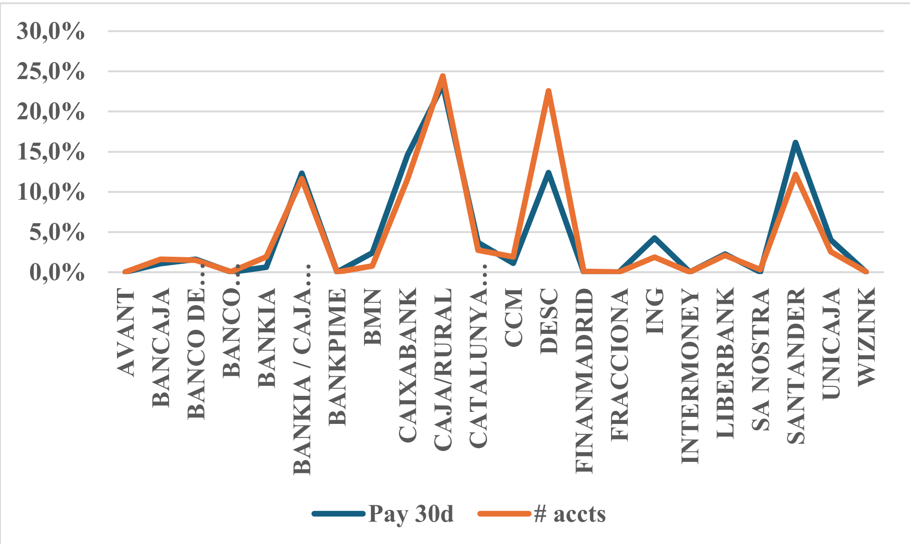


Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

In this case, we observe that there is certain correlation between the creditor grouped and the payment in the following 30 days (the distribution of accounts and payments is not similar for the different creditors grouped). In Individuals, the probability for a payment is higher in these creditors: Caixabank, Catalunya Caixa, ING, Santander and Unicaja. For the SMEs, the higher probability belongs to Caixabank, Santander, Unicaja and BMN. The information value (IV) of this variable is 0.16 for Individuals and 0.15 for SMEs.

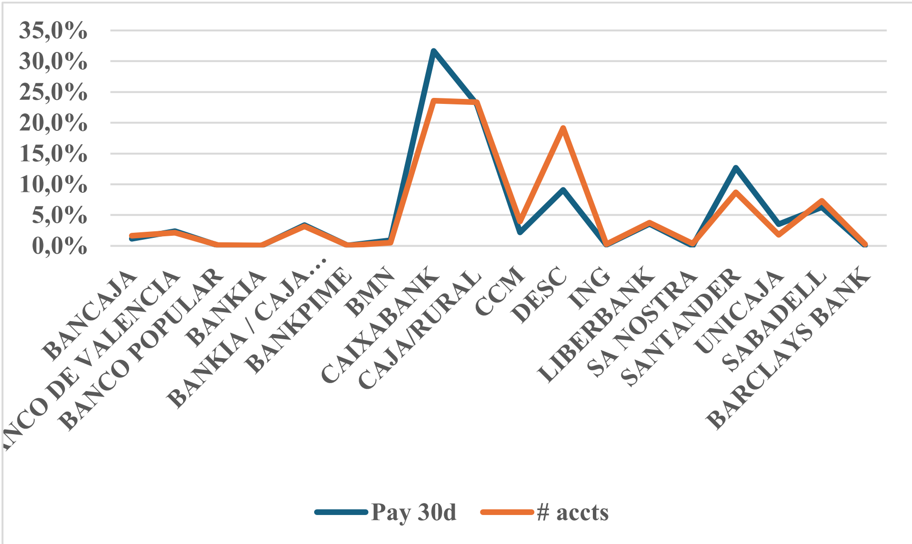
Figure 136 and Table 145. Bivariant analysis creditor grouped – payment 30 days for Individuals



Creditor grouped	Pay 30d	# accts
AVANT	0,0%	0,0%
BANCAJA	1,1%	1,6%
BANCO DE VALENCIA	1,6%	1,5%
BANCO POPULAR	0,0%	0,1%
BANKIA	0,6%	1,9%
BANKIA / CAJA MADRID	12,3%	11,6%
BANKPIME	0,0%	0,0%
BMN	2,4%	0,8%
CAIXABANK	14,6%	11,6%
CAJA/RURAL	23,3%	24,4%
CATALUNYA CAIXA	3,7%	2,7%
CCM	1,1%	1,9%
DESC	12,4%	22,6%
FINANMADRID	0,0%	0,1%
FRACCIONA	0,0%	0,0%
ING	4,2%	1,9%
INTERMONEY	0,0%	0,0%
LIBERBANK	2,3%	2,1%
SA NOSTRA	0,0%	0,3%
SANTANDER	16,2%	12,2%
UNICAJA	4,0%	2,6%
WIZINK	0,0%	0,0%
Total general	**100,0%**	**100,0%**

Source: Own elaboration

Figure 137 and Table 146. Bivariant analysis creditor grouped – payment 30 days for SMEs



Creditor grouped	Pay 30d	# accts
BANCAJA	1,1%	1,6%
BANCO DE VALENCIA	2,4%	2,1%
BANCO POPULAR	0,0%	0,1%
BANKIA	0,0%	0,1%
BANKIA / CAJA MADRID	3,3%	3,1%
BANKPIME	0,1%	0,1%
BMN	0,9%	0,5%
CAIXABANK	31,7%	23,6%
CAJA/RURAL	23,0%	23,4%
CCM	2,2%	3,9%
DESC	9,1%	19,1%
ING	0,1%	0,2%
LIBERBANK	3,6%	3,8%
SA NOSTRA	0,0%	0,4%
SANTANDER	12,7%	8,7%
UNICAJA	3,5%	1,8%
SABADELL	6,3%	7,3%
BARCLAYS BANK	0,0%	0,2%
Total general	**100,0%**	**100,0%**

Source: Own elaboration

3.9.- Variable "IN13_Nacimiento_Min" – Oldest date of birth or date of company creation of participants

Description

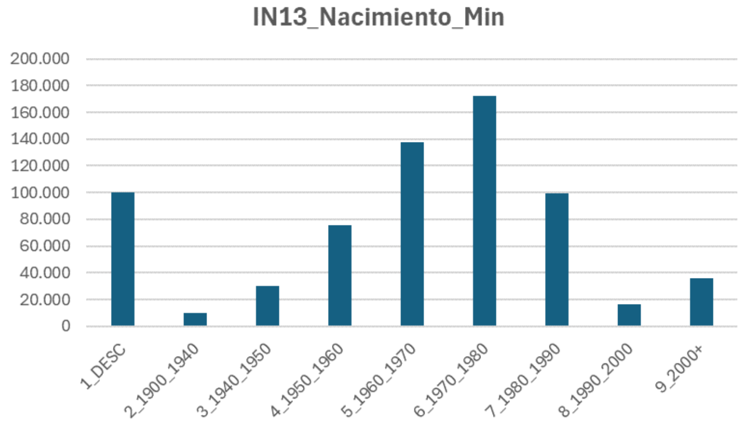
Oldest date of birth or date of the company creation for any of the participants in the debt or the company. This variable was included in the data set, as normally shows correlation with the probability to collect. Debt of older participants shows higher probability to collect, and younger participants shows lower probability. Finally, this variable was relevant in our modelling only for the SMEs Covid-19 model.

Individuals' details

High concentration in cases of debtors born between 1960 and 1980, 46% of the total accounts. Around 15% of cases have null or no information values.

Table 147 and Figure 138. Distribution of Individuals accounts by oldest participant date

IN13_Nacimiento_Min	Record_Count
1_DESC	99.948
2_1900_1940	9.638
3_1940_1950	29.920
4_1950_1960	75.295
5_1960_1970	137.357
6_1970_1980	172.067
7_1980_1990	99.415
8_1990_2000	16.632
9_2000+	35.756



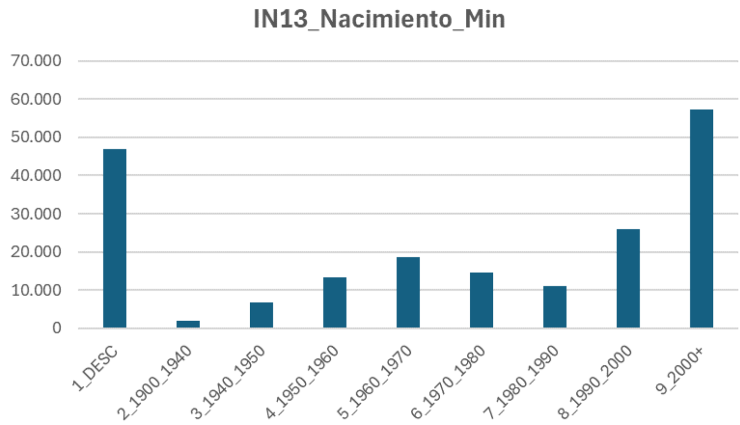
Source: Own elaboration

SMEs details

High concentration in cases of young companies or of recent creation, 29% of the total accounts are from companies of a creation after 2000. 24% of cases have null or no information values.

Table 148 and Figure 139. Distribution of SMEs accounts by oldest participant date

IN13_Nacimiento_Min	Record_Count
1_DESC	46.864
2_1900_1940	1.862
3_1940_1950	6.683
4_1950_1960	13.190
5_1960_1970	18.652
6_1970_1980	14.577
7_1980_1990	11.062
8_1990_2000	25.850
9_2000+	57.222



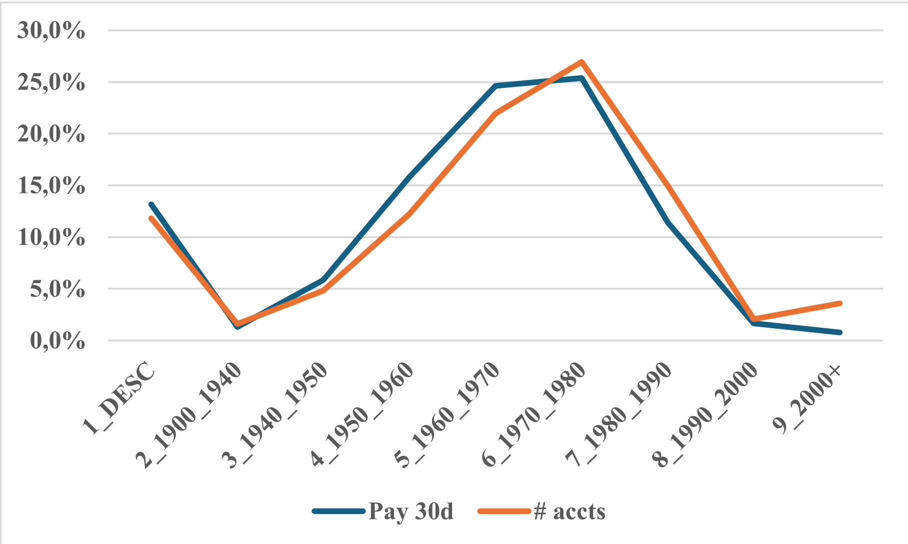
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

In this case, we observe that there is certain correlation between the oldest participant and the payment in the following 30 days (the distribution of accounts and payments is not similar for the different ranges of oldest participant). In Individuals, the probability for a payment is higher when the oldest participant is from the 1940-1950, 1950-1960 and 1960-1970 ranges. And in SMEs, this probability is higher for the ranges among 1900 and 1980. The information value (IV) of this variable is 0.07 for Individuals and 0.81 for SMEs.

Table 149 and Figure 140. Bivariant analysis oldest participant – payment 30 days for Individuals

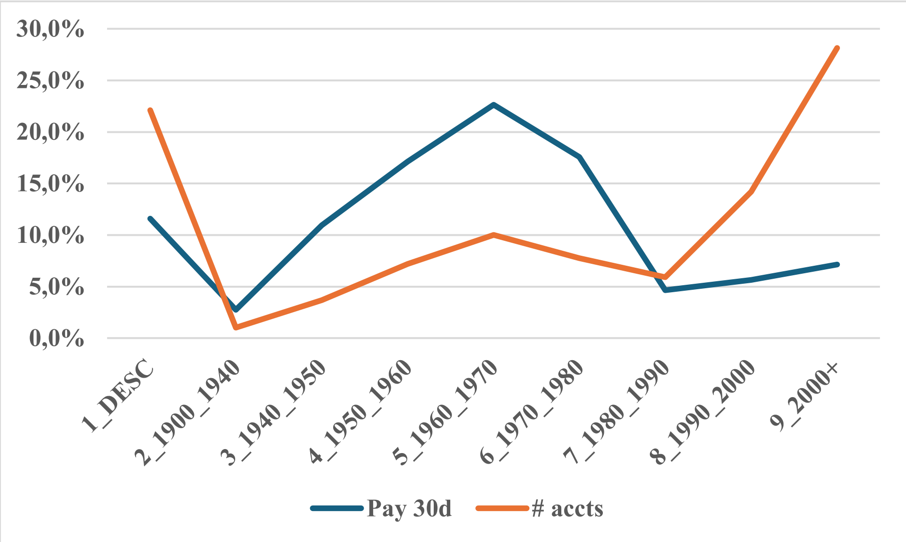
Oldest DOB	Pay 30d	# accts
1_DESC	13,2%	11,8%
2_1900_1940	1,3%	1,6%
3_1940_1950	5,8%	4,8%
4_1950_1960	15,8%	12,2%
5_1960_1970	24,6%	22,0%
6_1970_1980	25,4%	27,0%
7_1980_1990	11,4%	15,0%
8_1990_2000	1,6%	2,1%
9_2000+	0,8%	3,6%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 150 and Figure 141. Bivariant analysis oldest participant – payment 30 days for SMEs

Oldest DOB	Pay 30d	# accts
1_DESC	11,6%	22,1%
2_1900_1940	2,7%	1,0%
3_1940_1950	10,9%	3,7%
4_1950_1960	17,1%	7,2%
5_1960_1970	22,6%	10,0%
6_1970_1980	17,6%	7,8%
7_1980_1990	4,7%	5,9%
8_1990_2000	5,6%	14,2%
9_2000+	7,1%	28,2%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.10.- Variable "IN14_Nacimiento_Max" – Youngest date of birth or date of company creation of participants

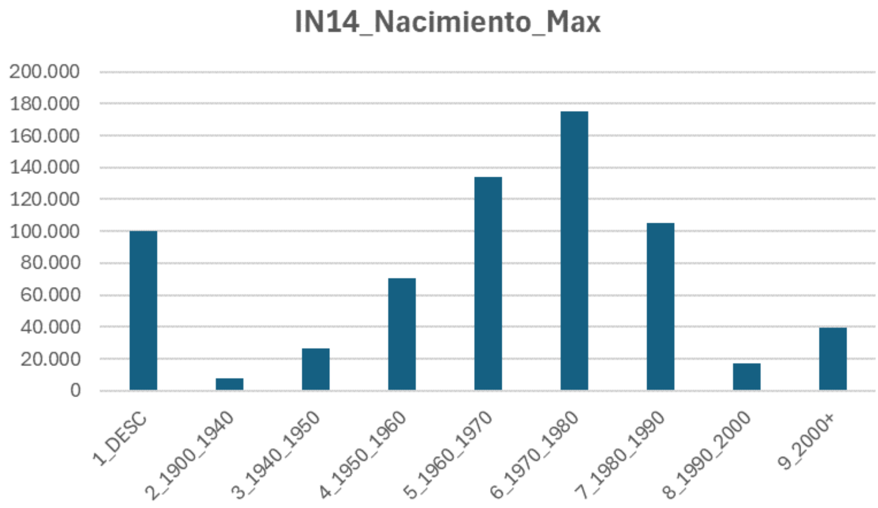
Description

Youngest date of birth or date of the company creation for any of the participants in the debt or the company. This variable was included in the data set, as normally shows correlation with the probability to collect. Debt of older participants shows higher probability to collect, and younger participants shows lower probability. Finally, this variable was relevant for the Individuals Pre Covid-19 and all the SMEs models.

Individuals' details

Very similar results than with the variable “IN13_Nacimiento_Min” – oldest date of birth or date of company creation of participants. High concentration in cases of debtors born between 1960 and 1980, 46% of the total accounts. Around 15% of cases have null or no information values.

Figure 142 and Table 151. Distribution of Individuals accounts by youngest participant date



IN14_Nacimiento_Max	Record_Count
1_DESC	99.948
2_1900_1940	7.588
3_1940_1950	26.798
4_1950_1960	70.505
5_1960_1970	134.419
6_1970_1980	175.084
7_1980_1990	104.897
8_1990_2000	17.342
9_2000+	39.447

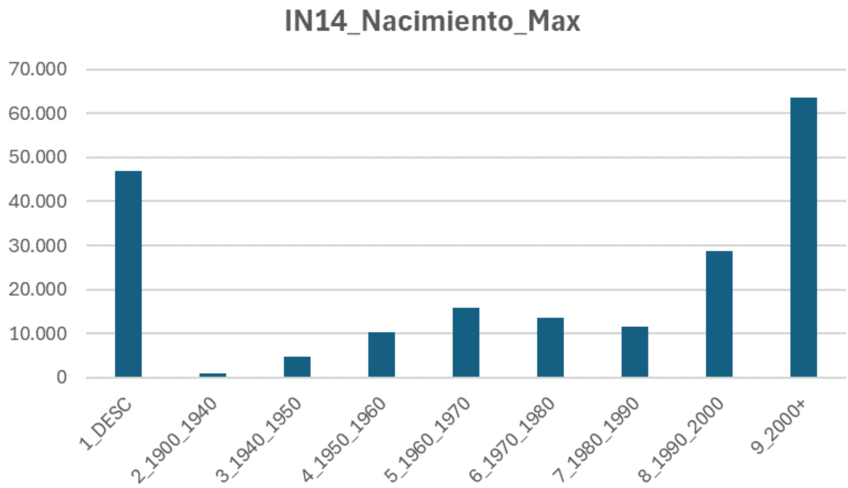
Source: Own elaboration

SMEs details

Very similar results than with the variable “IN13_Nacimiento_Min” – oldest date of birth or date of company creation of participants. High concentration in cases of young companies or of recent creation, 32% of the total accounts are from companies of a creation after 2000. 24% of cases have null or no information values.

Table 152 and Figure 143. Distribution of SMEs accounts by youngest participant date

IN14_Nacimiento_Max	Record_Count
1_DESC	46.864
2_1900_1940	937
3_1940_1950	4.641
4_1950_1960	10.182
5_1960_1970	15.784
6_1970_1980	13.629
7_1980_1990	11.566
8_1990_2000	28.703
9_2000+	63.656



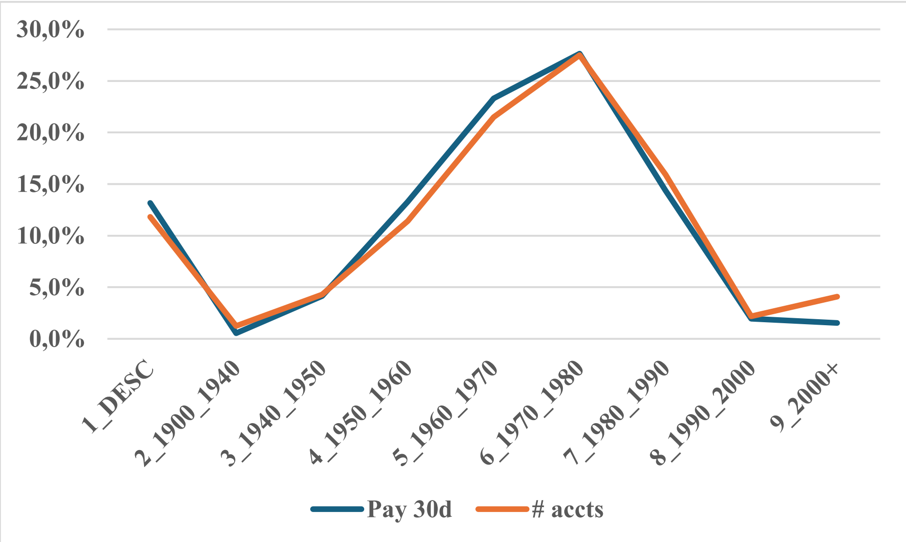
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

In this case, we observe a very different situation in terms of correlation between the youngest participant and the payment in the following 30 days (comparison of the distribution of accounts and payments for the different ranges of youngest participant). In Individuals, there is limited correlation. While for the SMEs we observe an important correlation, being the probability higher in the ranges between 1900 and 1980. The information value (IV) of this variable is 0.04 for Individuals and 0.52 for SMEs.

Table 153 and Figure 144. Bivariant analysis youngest participant – payment 30 days for Individuals

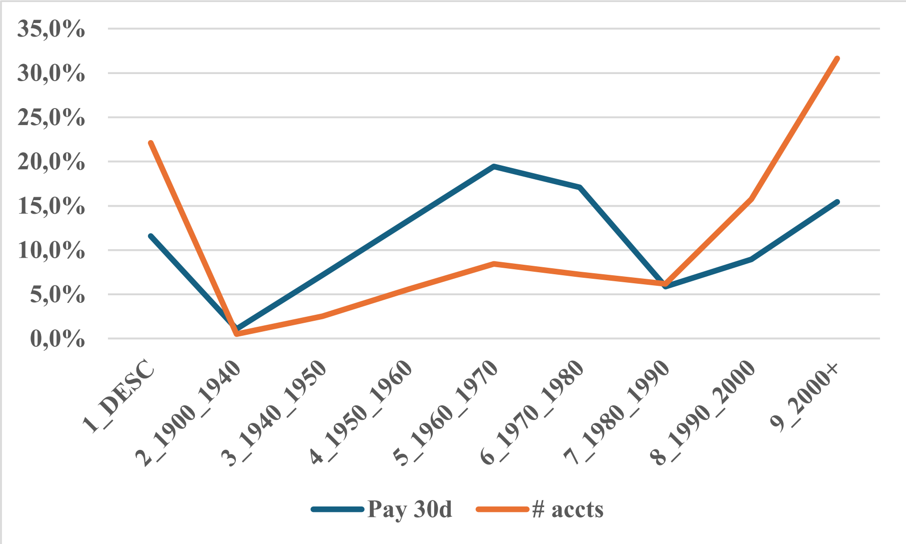
Youngest DOB	Pay 30d	# accts
1_DESC	13,2%	11,8%
2_1900_1940	0,6%	1,3%
3_1940_1950	4,2%	4,3%
4_1950_1960	13,3%	11,4%
5_1960_1970	23,3%	21,5%
6_1970_1980	27,6%	27,5%
7_1980_1990	14,4%	15,9%
8_1990_2000	1,9%	2,2%
9_2000+	1,5%	4,1%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 154 and Figure 145. Bivariant analysis youngest participant – payment 30 days for SMEs

Youngest DOB	Pay 30d	# accts
1_DESC	11,6%	22,1%
2_1900_1940	1,1%	0,5%
3_1940_1950	7,2%	2,5%
4_1950_1960	13,3%	5,6%
5_1960_1970	19,5%	8,4%
6_1970_1980	17,1%	7,2%
7_1980_1990	5,9%	6,2%
8_1990_2000	9,0%	15,8%
9_2000+	15,5%	31,6%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.11.- Variable "IN15_NUM_TEL_FIJO_Sum" – Number of fixed line phone numbers

Description

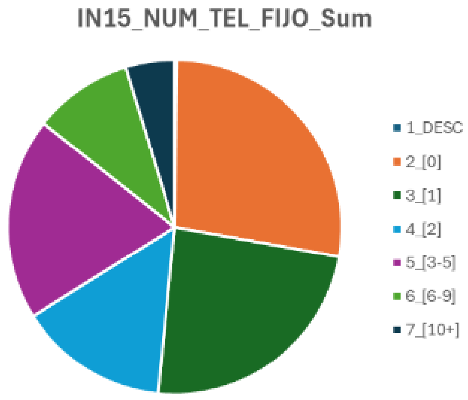
This variable represents the number of fixed line phone numbers available for the account, grouped in several ranges. This variable was included in the data set, as normally shows correlation with the probability to collect. The existence of phone numbers and the number of phone numbers have a positive correlation with the probability to collect. In our modelling, this variable was relevant for the Individuals Global, Individuals Pre Covid-19 and SMEs Post Covid-19 models.

Individuals' details

38% of the accounts have one or two available fixed line phone numbers, and approximately 5% of the accounts have ten or more numbers. Around 28% of cases have none fixed line phone number available.

Table 155 and Figure 146. Distribution of Individuals accounts by number of fixed line phone numbers

IN15_NUM_TEL_FIJO_Sum	Record_Count
1_DESC	1.476
2_[0]	186.621
3_[1]	160.804
4_[2]	97.800
5_[3-5]	132.397
6_[6-9]	64.835
7_[10+]	32.095



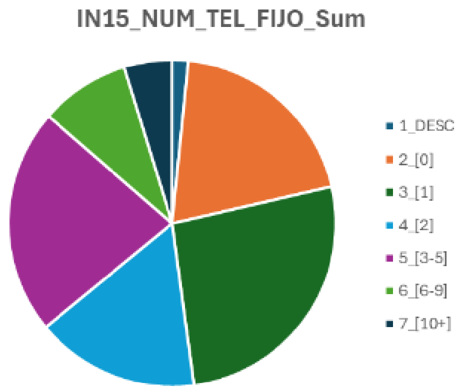
Source: Own elaboration

SMEs details

43% of the accounts have one or two available fixed line phone numbers, and approximately 5% of the accounts have ten or more numbers. Around 21% of cases have none fixed line phone number available.

Table 156 and Figure 147. Distribution of SMEs accounts by number of fixed line phone numbers

IN15_NUM_TEL_FIJO_Sum	Record_Count
1_DESC	3.124
2_[0]	38.775
3_[1]	51.904
4_[2]	31.730
5_[3-5]	43.746
6_[6-9]	17.399
7_[10+]	9.284



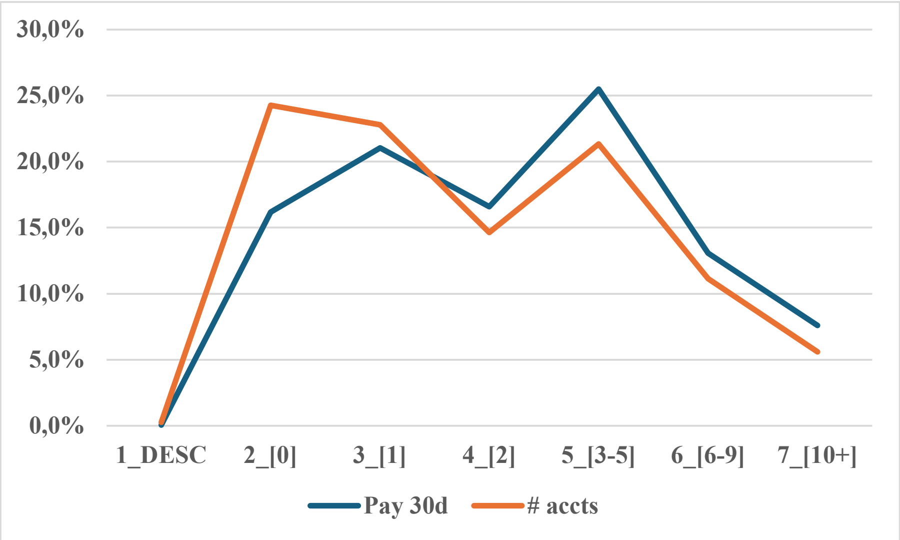
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

In this case, we observe that there is certain correlation between the number of fixed line phone numbers and the payment in the following 30 days (the distribution of accounts and payments is not similar for the different ranges of number of fixed line phone numbers). The probability for a payment is higher when the number of fixed line phone numbers is higher. In Individuals we observe this with 2 or more, and for SMEs with 3 or more. The information value (IV) of this variable is 0.06 for Individuals and 0.19 for SMEs.

Table 157 and Figure 148. Bivariant analysis number of fixed line phone numbers – payment 30 days for Individuals

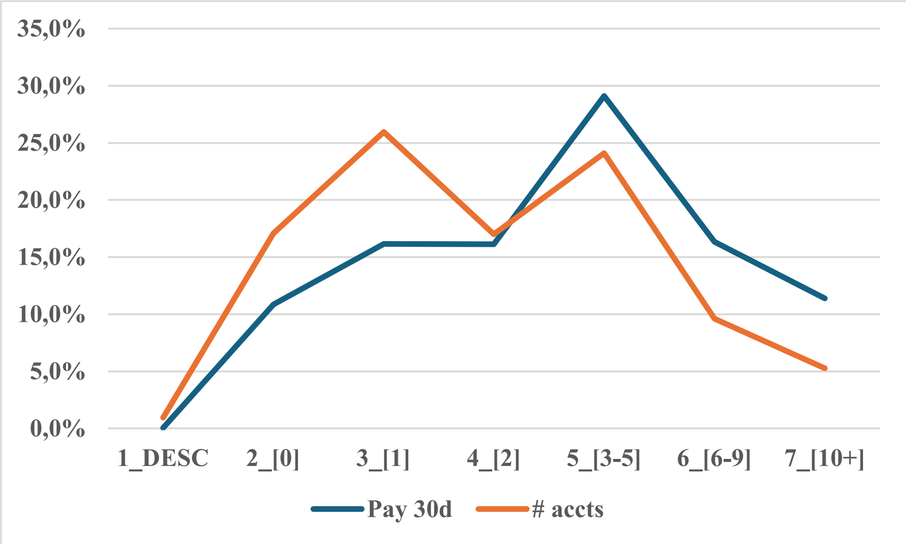
# of fixed line phone numbers	Pay 30d	# accts
1_DESC	0,0%	0,2%
2_[0]	16,2%	24,3%
3_[1]	21,0%	22,8%
4_[2]	16,6%	14,6%
5_[3-5]	25,5%	21,3%
6_[6-9]	13,1%	11,1%
7_[10+]	7,6%	5,6%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 158 and Figure 149. Bivariant analysis number of fixed line phone numbers – payment 30 days for SMEs

# of fixed line phone numbers	Pay 30d	# accts
1_DESC	0,0%	1,0%
2_[0]	10,8%	17,1%
3_[1]	16,1%	26,0%
4_[2]	16,1%	17,0%
5_[3-5]	29,1%	24,1%
6_[6-9]	16,3%	9,6%
7_[10+]	11,4%	5,3%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.12.- Variable “IN16_NUM_TEL_MOVIL_Sum” – Number of mobile phone numbers

Description

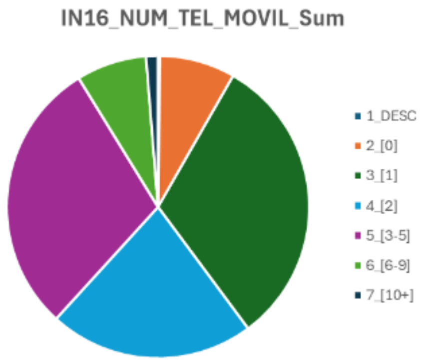
This variable represents the number of mobile phone numbers available for the account, grouped in several ranges. This variable was included in the data set, as normally shows correlation with the probability to collect. The existence of phone numbers and the number of phone numbers have a positive correlation with the probability to collect. Within our modelling, this variable was relevant in the Individuals Global, SMEs Global, SMEs Pre Covid-19 and SMEs Post Covid-19 models.

Individuals’ details

54% of the accounts have one or two available mobile phone numbers, and approximately 1% of the accounts have ten or more numbers. Around 8% of cases have none mobile phone number available.

Table 159 and Figure 150. Distribution of Individuals accounts by number of mobile phone numbers

IN16_NUM_TEL_MOVIL_Sum	Record_Count
1_DESC	1.476
2_[0]	54.523
3_[1]	213.372
4_[2]	148.210
5_[3-5]	199.095
6_[6-9]	50.773
7_[10+]	8.579



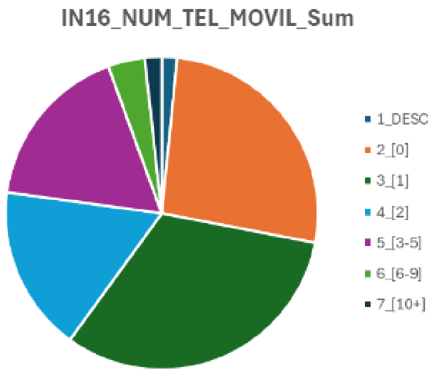
Source: Own elaboration

SMEs details

50% of the accounts have one or two available mobile phone numbers, and approximately 2% of the accounts have ten or more numbers. Around 28% of cases have none mobile phone number available, what is more common in the companies' segment (compared with the Individuals figures).

Table 160 and Figure 151. Distribution of SMEs accounts by number of mobile phone numbers

IN16_NUM_TEL_MOVIL_Sum	Record_Count
1_DESC	3.124
2_[0]	51.799
3_[1]	62.671
4_[2]	33.582
5_[3-5]	33.900
6_[6-9]	7.472
7_[10+]	3.414



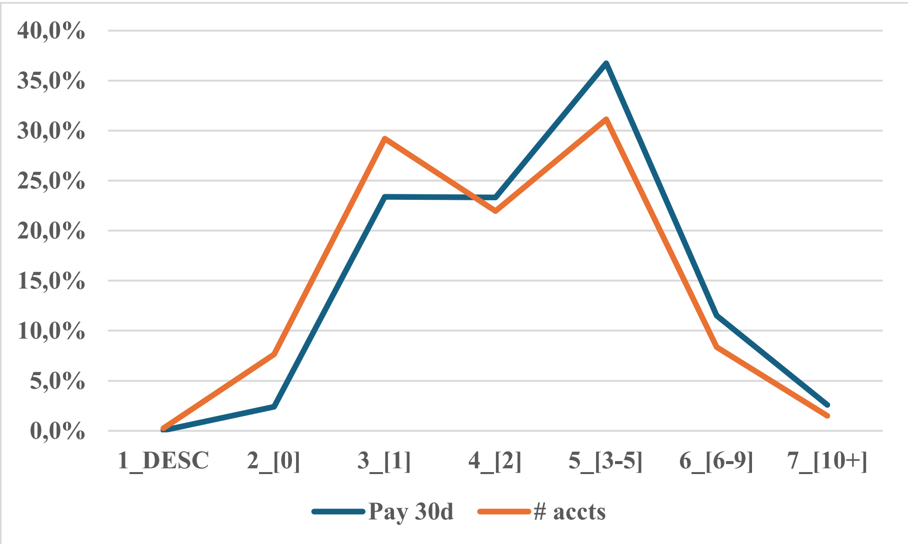
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

In this case, we observe that there is certain correlation between the number of mobile phone numbers and the payment in the following 30 days (the distribution of accounts and payments is not similar for the different ranges of the number of mobile phone numbers). Both for Individuals and SMEs, the probability for a payment is higher when the number of mobile phone numbers is 2 or more. The information value (IV) of this variable is 0.10 for Individuals and 0.30 for SMEs.

Table 161 and Figure 152. Bivariant analysis number of mobile phone numbers – payment 30 days for Individuals

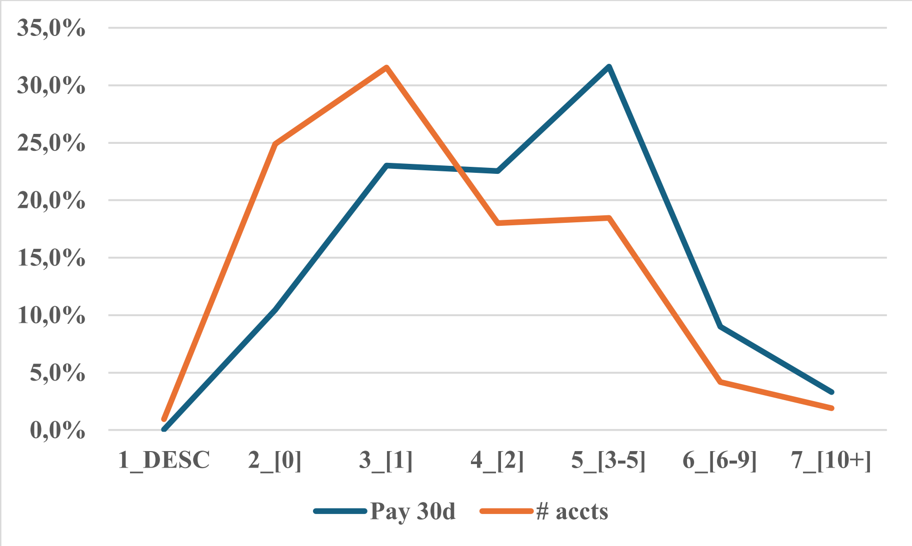
# of mobile phone numbers	Pay 30d	# accts
1_DESC	0,0%	0,2%
2_[0]	2,4%	7,6%
3_[1]	23,4%	29,2%
4_[2]	23,3%	21,9%
5_[3-5]	36,7%	31,1%
6_[6-9]	11,5%	8,4%
7_[10+]	2,6%	1,5%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 162 and Figure 153. Bivariant analysis number of mobile phone numbers – payment 30 days for SMEs

# of mobile phone numbers	Pay 30d	# accts
1_DESC	0,0%	1,0%
2_[0]	10,5%	24,9%
3_[1]	23,0%	31,6%
4_[2]	22,5%	18,0%
5_[3-5]	31,6%	18,5%
6_[6-9]	9,0%	4,2%
7_[10+]	3,3%	1,9%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.13.- Variable "IN17_NE_EXT" – Foreign nationality flag

Description

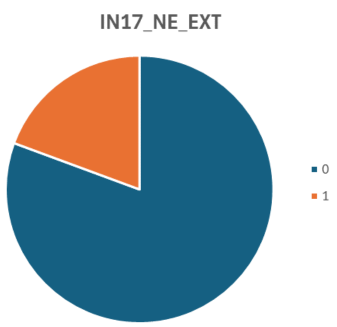
This variable is linked with the nationality and origin of the account, and it represents a flag if any of the participants have a foreign nationality or not. This variable was included in the data set, as normally shows correlation with the probability to collect. In general, the foreign debtors have a lower probability to collect, as it is more difficult to trace them. In our modelling, this variable was relevant only for the Individuals Pre Covid-19 model.

Individuals' details

19% of the accounts have a foreign nationality flag.

Table 163 and Figure 154. Distribution of Individuals accounts by foreign nationality flag

IN17_NE_EXT	Record_Count
0	545.081
1	130.947



Source: Own elaboration

SMEs details

Just 1% of the accounts have a foreign nationality flag.

Table 164 and Figure 155. Distribution of SMEs accounts by foreign nationality flag

IN17_NE_EXT	Record_Count
0	193.439
1	2.523

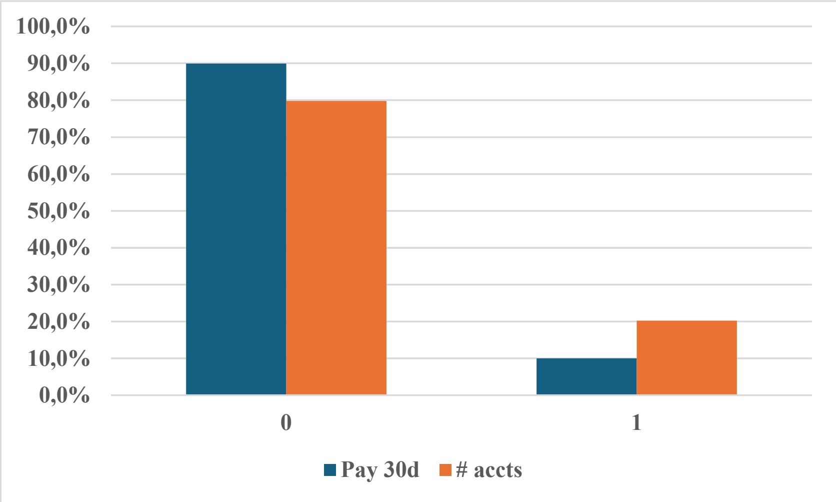
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

In this case, we observe that there is certain correlation between the foreign nationality flag and the payment in the following 30 days (the distribution of accounts and payments is not similar for the different foreign nationality flags). In Individuals, the probability for a payment is lower for the cases positive in the foreign nationality flag. But in SMEs we do not observe a significant difference in behaviour. The information value (IV) of this variable is 0.08 for Individuals and 0.00 for SMEs.

Table 165 and Figure 156. Bivariant analysis foreign nationality flag – payment 30 days for Individuals

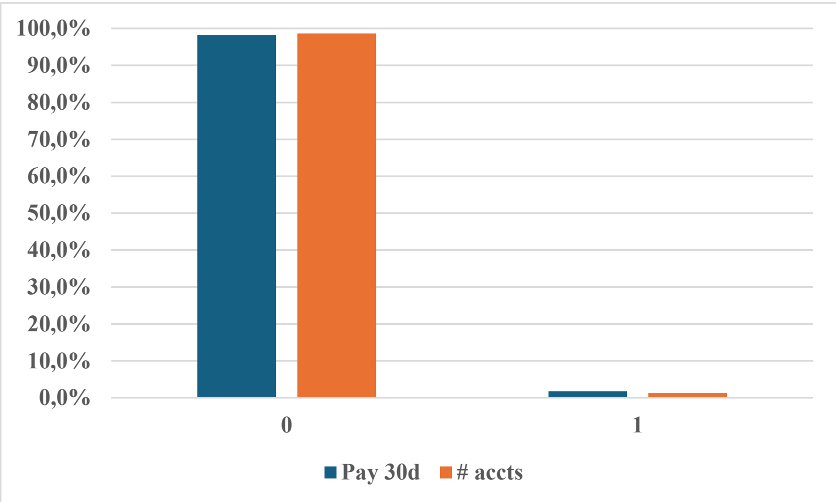
Foreign nationality flag	Pay 30d	# accts
0	90,0%	79,8%
1	10,0%	20,2%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 166 and Figure 157. Bivariant analysis foreign nationality flag – payment 30 days for SMEs

Foreign nationality flag	Pay 30d	# accts
0	98,2%	98,7%
1	1,8%	1,3%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.14.- Variable "IN18_NE_NAC" – National nationality flag

Description

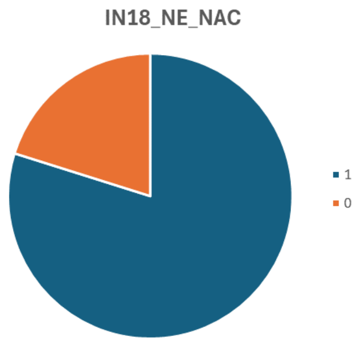
This variable is linked with the nationality and origin of the account, and it represents a flag if any of the participants have a national nationality or not. This variable was included in the data set, as normally shows correlation with the probability to collect. In general, the national debtors have a higher probability to collect compared with the foreign debtors. In our analysis, this variable was relevant for the Individuals Global, Individuals Pre Covid-19 and Individuals Post Covid-19 models.

Individuals' details

80% of the accounts have a national nationality flag.

Table 167 and Figure 158. Distribution of Individuals accounts by national nationality flag

IN18_NE_NAC	Record_Count
1	539.771
0	136.257



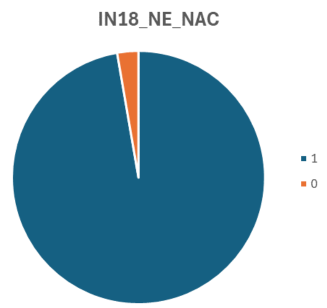
Source: Own elaboration

SMEs details

97% of the accounts have a national nationality flag.

Table 168 and Figure 159. Distribution of SMEs accounts by national nationality flag

IN18_NE_NAC	Record_Count
1	190.660
0	5.302



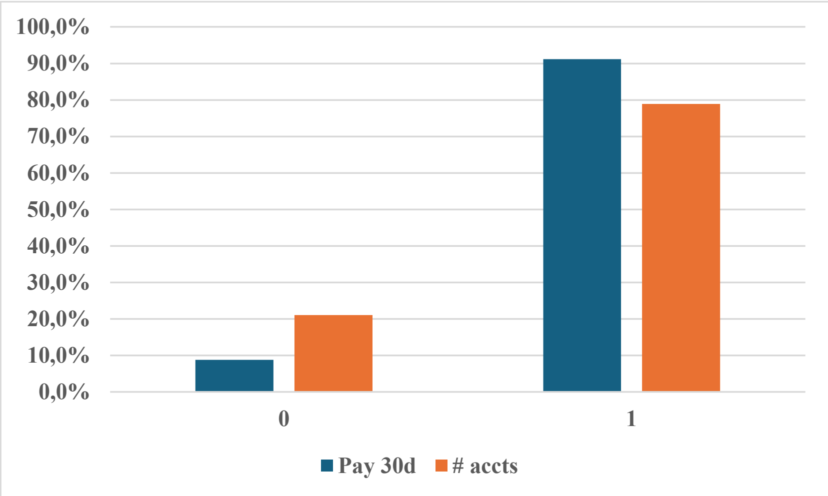
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

In this case, we observe that there is certain correlation between the national nationality flag and the payment in the following 30 days (the distribution of accounts and payments is not similar for the different national nationality flags). In Individuals, the probability for a payment is higher for the cases positive in the national nationality flag. But in SMEs there is not differences in the behaviour observed. The information value (IV) of this variable is 0.12 for Individuals and 0.00 for SMEs.

Table 169 and Figure 160. Bivariant analysis national nationality flag – payment 30 days for Individuals

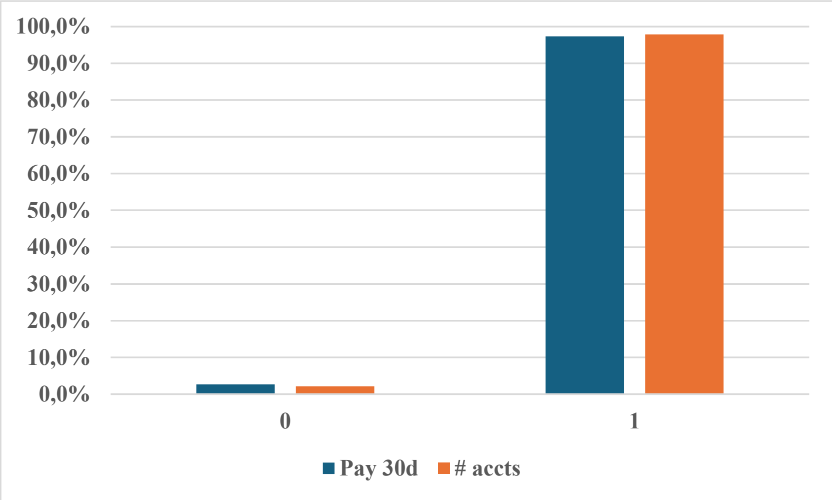
National nationality flag	Pay 30d	# accts
0	8,8%	21,1%
1	91,2%	78,9%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 170 and Figure 161. Bivariant analysis national nationality flag – payment 30 days for SMEs

National nationality flag	Pay 30d	# accts
0	2,6%	2,1%
1	97,4%	97,9%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.15.- Variable "IN19_NE_OTH" – Other nationality flag

Description

This variable is linked with the nationality and origin of the account, and it represents a flag if any of the participants have an unknown nationality or not; and it was included in the data set, as normally shows correlation with the probability to collect. In general, the national debtors have a higher probability to collect and the foreign debtors have it lower. Finally, this variable was relevant in our modelling only for the Individuals Post Covid-19 model.

Individuals' details

Just 1% of the accounts have an unknown nationality flag.

Table 171 and Figure 162. Distribution of Individuals accounts by other nationality flag

IN19_NE_OTH	Record_Count
0	667.132
1	8.896

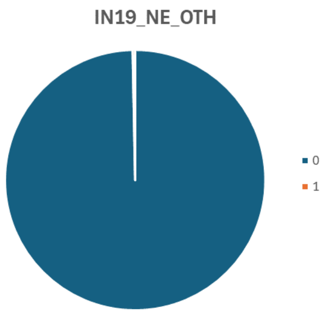
Source: Own elaboration

SMEs details

Only 0,3% of the accounts have an unknown nationality flag.

Table 172 and Figure 163. Distribution of SMEs accounts by other nationality flag

IN19_NE_OTH	Record_Count
0	195.362
1	600



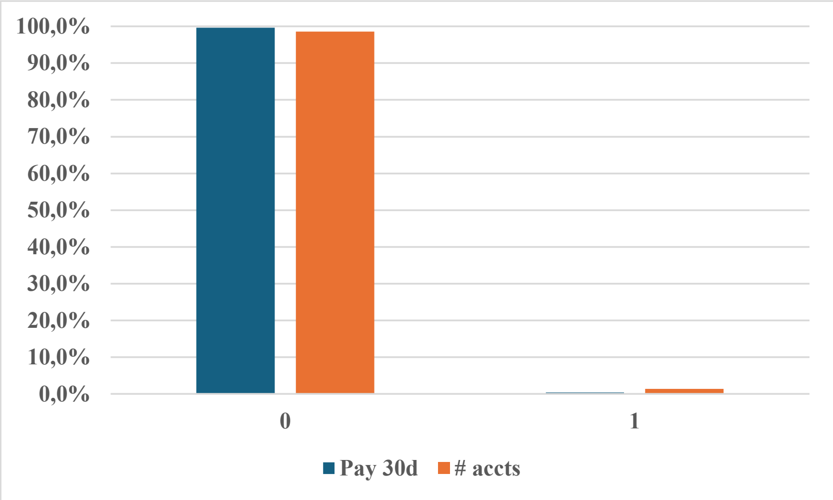
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

In this case, we observe that there is certain, but limited, correlation between the other nationality flag and the payment in the following 30 days (the distribution of accounts and payments is not similar for the different other nationality flags). The probability for a payment is lower for the cases positive in the other nationality flag, both in Individuals and SMEs. The information value (IV) of this variable is 0.01 for Individuals and 0.02 for SMEs.

Table 173 and Figure 164. Bivariant analysis other nationality flag – payment 30 days for Individuals

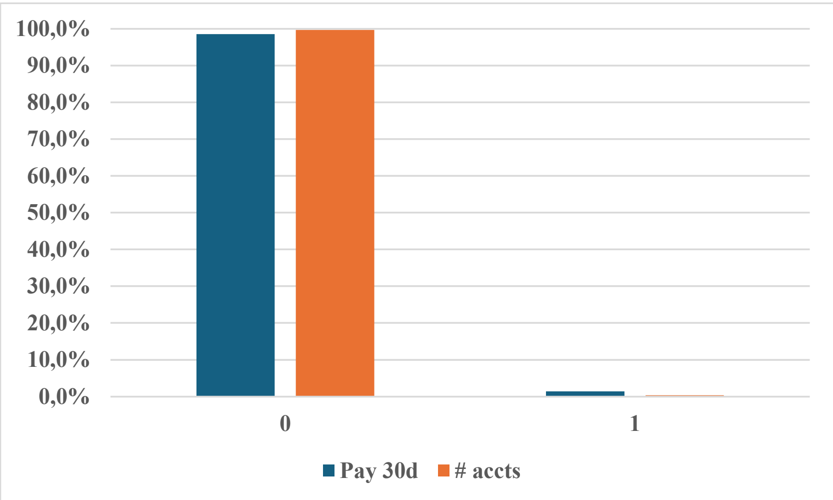
Other nationality flag	Pay 30d	# accts
0	99,6%	98,6%
1	0,4%	1,4%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 174 and Figure 165. Bivariant analysis other nationality flag – payment 30 days for SMEs

Other nationality flag	Pay 30d	# accts
0	98,6%	99,7%
1	1,4%	0,3%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.16.- Variable "IN20_Provincia" – Province

Description

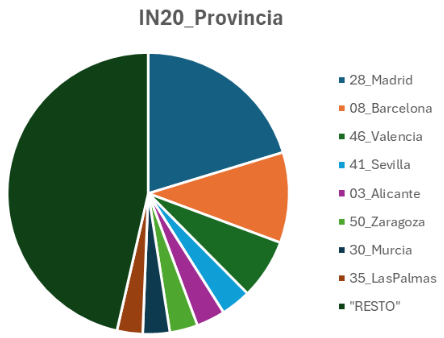
This variable identifies the province where the debt has its origin; and it was included in the data set, as normally shows correlation with the probability to collect. The region where the debt was originated have a clear correlation with the probability to collect, as the demographic and macroeconomics factors are very different among the regions, and these factors directly influence in the probability to collect. Certain cities, provinces and regions have a higher probability to collect than others. In our modelling, this variable was relevant only for the SMEs Covid-19 model.

Individuals' details

The accounts are very well spread around the Spanish national territory. Madrid and Barcelona concentrate 31% of the total accounts. There is as well an important concentration on the Mediterranean Coast (Valencia, Alicante and Murcia), with around 13% of the accounts.

Table 175 and Figure 166. Distribution of Individuals accounts by province

IN20_Provincia	Record_Count
28_Madrid	137147
08_Barcelona	70264
46_Valencia	46631
41_Sevilla	23369
03_Alicante	22264
50_Zaragoza	21710
30_Murcia	20904
35_LasPalmas	19939
"RESTO"	313800



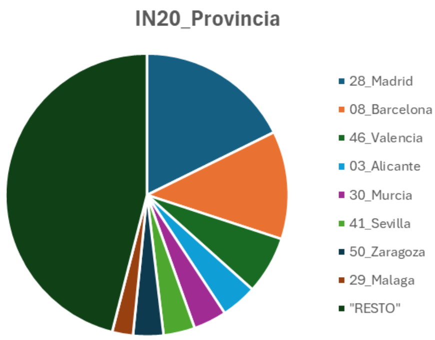
Source: Own elaboration

SMEs details

The accounts are very well spread around the Spanish national territory. Madrid and Barcelona concentrate 30% of the total accounts. There is as well an important concentration on the Mediterranean Coast (Valencia, Alicante and Murcia), with around 15% of the accounts.

Table 176 and Figure 167. Distribution of SMEs accounts by province

IN20_Provincia	Record_Count
28_Madrid	34.654
08_Barcelona	24.218
46_Valencia	12.935
03_Alicante	8.059
30_Murcia	7.408
41_Sevilla	7.039
50_Zaragoza	6.777
29_Malaga	4.674
33_Asturias	4.374
45_Toledo	4.270

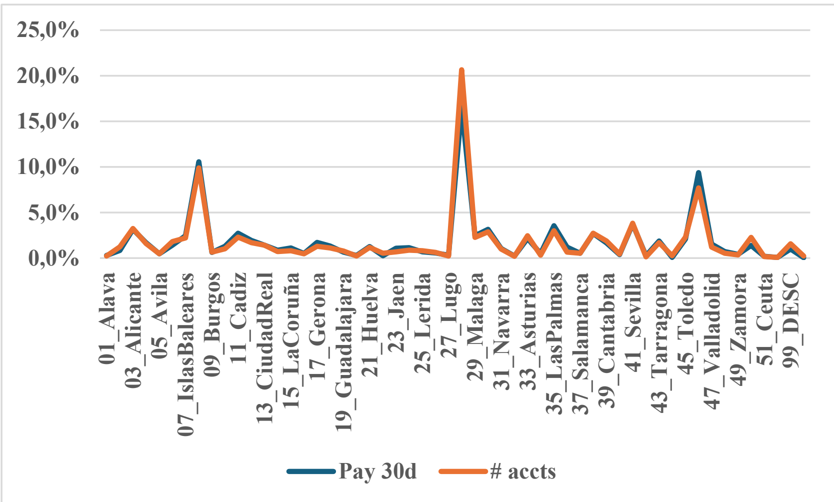


Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

In this case, we observe that there is certain correlation between the province and the payment in the following 30 days (the distribution of accounts and payments is not similar for the different provinces). The probability for a payment varies widely with the province. The information value (IV) of this variable is 0.05 for Individuals and 0.10 for SMEs.

Figure 168. Bivariant analysis province – payment 30 days for Individuals



Source: Own elaboration

Table 177. Bivariant analysis province – payment 30 days for Individuals

Province	Pay 30d	# accts
01_Alava	0,3%	0,3%
02_Albacete	0,9%	1,3%
03_Alicante	3,2%	3,3%
04_Almeria	1,7%	1,6%
05_Avila	0,5%	0,5%
06_Badajoz	1,4%	1,8%
07_IslasBaleares	2,5%	2,2%
08_Barcelona	10,6%	9,9%
09_Burgos	0,6%	0,7%
10_Caceres	1,3%	1,0%
11_Cadiz	2,7%	2,3%
12_Castellon	1,9%	1,7%
13_CiudadReal	1,4%	1,4%
14_Cordoba	0,9%	0,7%
15_LaCoruña	1,1%	0,8%
16_Cuenca	0,5%	0,5%
17_Gerona	1,7%	1,3%
18_Granada	1,3%	1,1%
19_Guadalajara	0,7%	0,8%
20_Guipuzcoa	0,3%	0,3%
21_Huelva	1,3%	1,2%
22_Huesca	0,3%	0,6%
23_Jaen	1,1%	0,7%
24_Leon	1,1%	0,9%
25_Lerida	0,7%	0,8%
26_LaRioja	0,6%	0,6%
27_Lugo	0,3%	0,2%
28_Madrid	16,6%	20,6%
29_Malaga	2,5%	2,3%
30_Murcia	3,2%	2,9%
31_Navarra	1,1%	1,0%
32_Orense	0,2%	0,2%
33_Asturias	2,2%	2,5%
34_Palencia	0,5%	0,3%
35_LasPalmas	3,6%	3,0%
36_Pontevedra	1,2%	0,7%
37_Salamanca	0,5%	0,6%
38_Tenerife	2,7%	2,7%
39_Cantabria	1,7%	1,9%
40_Segovia	0,4%	0,5%
41_Sevilla	3,7%	3,8%
42_Soria	0,3%	0,2%
43_Tarragona	1,9%	1,7%
44_Teruel	0,1%	0,3%
45_Toledo	2,1%	2,3%
46_Valencia	9,4%	7,7%
47_Valladolid	1,5%	1,2%
48_Vizcaya	0,7%	0,5%
49_Zamora	0,4%	0,4%
50_Zaragoza	1,4%	2,3%
51_Ceuta	0,2%	0,2%
52_Melilla	0,1%	0,1%
99_DESC	1,0%	1,6%
(en blanco)	0,0%	0,2%
Total general	**100,0%**	**100,0%**

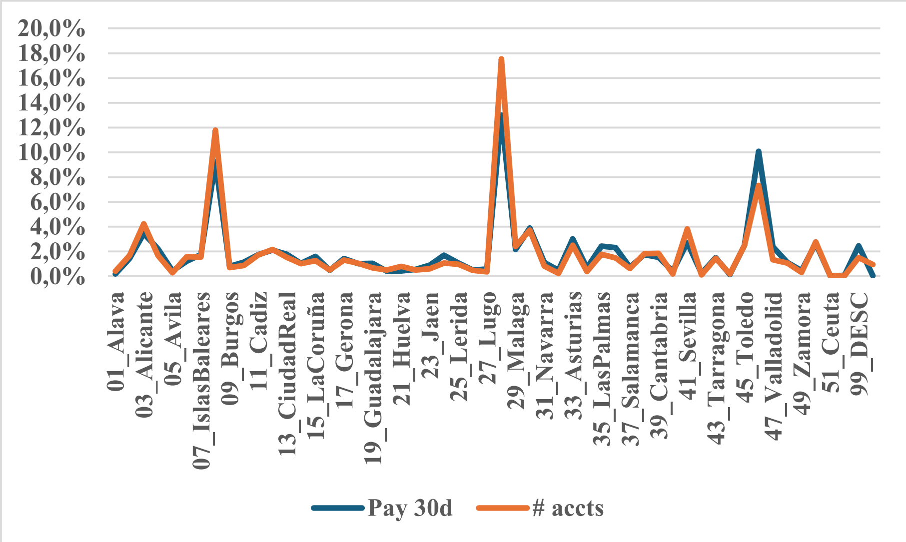
Source: Own elaboration

Table 178. Bivariant analysis province – payment 30 days for SMEs

Province	Pay 30d	# accts
01_Alava	0,2%	0,4%
02_Albacete	1,4%	1,7%
03_Alicante	3,4%	4,2%
04_Almeria	2,2%	1,7%
05_Avila	0,4%	0,3%
06_Badajoz	1,2%	1,6%
07_IslasBaleares	1,7%	1,5%
08_Barcelona	9,2%	11,8%
09_Burgos	0,8%	0,7%
10_Caceres	1,1%	0,9%
11_Cadiz	1,7%	1,7%
12_Castellon	2,1%	2,2%
13_CiudadReal	1,8%	1,5%
14_Cordoba	1,0%	1,0%
15_LaCoruña	1,6%	1,3%
16_Cuenca	0,5%	0,5%
17_Gerona	1,4%	1,3%
18_Granada	1,0%	1,0%
19_Guadalajara	1,0%	0,7%
20_Guipuzcoa	0,4%	0,5%
21_Huelva	0,4%	0,8%
22_Huesca	0,5%	0,5%
23_Jaen	0,9%	0,6%
24_Leon	1,7%	1,1%
25_Lerida	1,1%	1,0%
26_LaRioja	0,5%	0,5%
27_Lugo	0,6%	0,4%
28_Madrid	13,0%	17,5%
29_Malaga	2,2%	2,4%
30_Murcia	3,9%	3,7%
31_Navarra	1,1%	0,8%
32_Orense	0,5%	0,2%
33_Asturias	3,0%	2,5%
34_Palencia	0,7%	0,4%
35_LasPalmas	2,4%	1,8%
36_Pontevedra	2,3%	1,5%
37_Salamanca	0,7%	0,6%
38_Tenerife	1,8%	1,8%
39_Cantabria	1,5%	1,8%
40_Segovia	0,4%	0,2%
41_Sevilla	2,7%	3,8%
42_Soria	0,2%	0,1%
43_Tarragona	1,5%	1,5%
44_Teruel	0,1%	0,2%
45_Toledo	2,5%	2,5%
46_Valencia	10,1%	7,3%
47_Valladolid	2,4%	1,3%
48_Vizcaya	1,1%	1,1%
49_Zamora	0,5%	0,3%
50_Zaragoza	2,7%	2,8%
51_Ceuta	0,0%	0,0%
52_Melilla	0,1%	0,0%
99_DESC	2,5%	1,5%
(en blanco)	0,0%	1,0%
Total general	**100,0%**	**100,0%**

Source: Own elaboration

Figure 169. Bivariant analysis province – payment 30 days for SMEs



Source: Own elaboration

3.17.- Variable “IN21_TP_TIT” – Holder or active company flag

Description

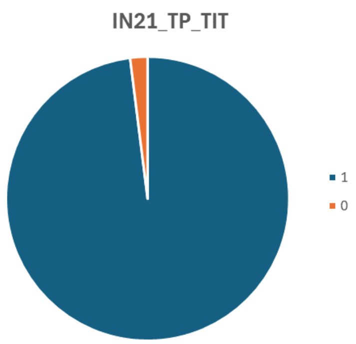
This variable identifies if the account counts with any holder or the company is active; and it was included in the data set, as normally shows correlation with the probability to collect. If the account has a holder or the company is active the probability to collect is higher than the opposite. Despite this typical behaviour, in our modelling this variable was not relevant for any of the models built.

Individuals’ details

More than 98% of the accounts have an active holder.

Table 179 and Figure 170. Distribution of Individuals accounts by holder or active company flag

IN21_TP_TIT	Record_Count
1	662.623
0	13.405



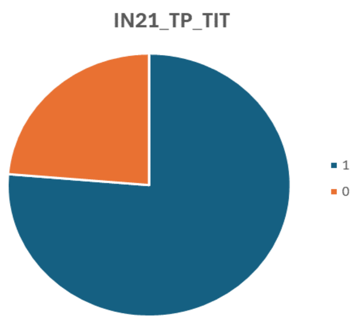
Source: Own elaboration

SMEs details

Around 75% of the accounts have an active holder.

Table 180 and Figure 171. Distribution of SMEs accounts holder or active company flag

IN21_TP_TIT	Record_Count
1	149.578
0	46.384



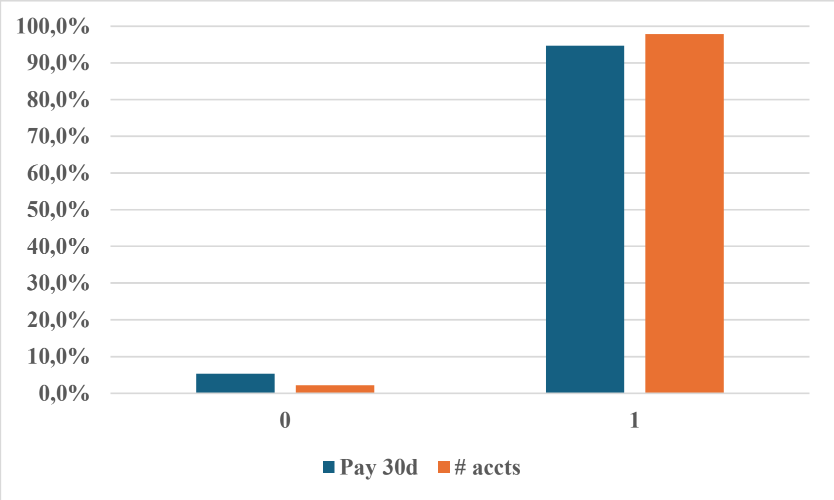
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

In this case, we observe that there is certain correlation between the holder or active company flag and the payment in the following 30 days (the distribution of accounts and payments is not similar for the different holder or active company flags). The probability for a payment is higher when the holder or active company flag is negative. The relevance of this flag is much higher in SMEs than in Individuals. The information value (IV) of this variable is 0.03 for Individuals and 0.56 for SMEs.

Table 181 and Figure 172. Bivariant analysis holder or active company flag – payment 30 days for Individuals

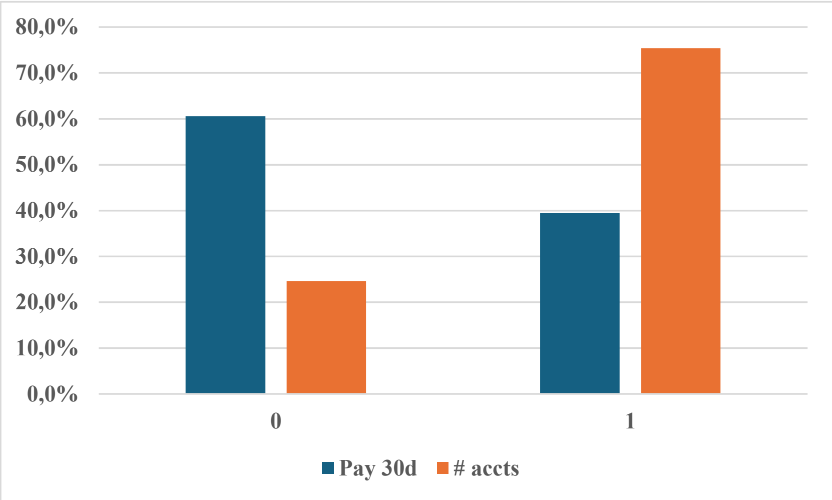
Holder or active company flag	Pay 30d	# accts
0	5,3%	2,2%
1	94,7%	97,8%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 182 and Figure 173. Bivariant analysis holder or active company flag – payment 30 days for SMEs

Holder or active company flag	Pay 30d	# accts
0	60,6%	24,6%
1	39,4%	75,4%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.18.- Variable "IN23_TP_OTH" – Other participant or relationship flag

Description

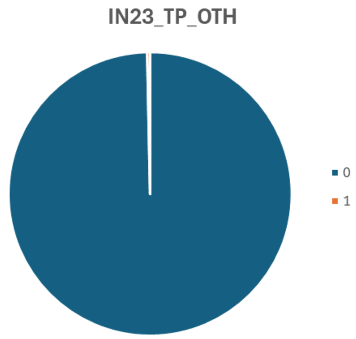
This variable identifies if the account counts with other participant or relationship different to holder or guarantor. This variable was included in the data set, as normally shows correlation with the probability to collect. If the account has other participants or relationships, the probability to collect is higher than the opposite. Within our modelling, this variable was not relevant in any of the models developed.

Individuals' details

Just 0,3% of the accounts have other participant or relationship.

Table 183 and Figure 174. Distribution of Individuals accounts by other participant or relationship flag

IN23_TP_OTH	Record_Count
0	673.810
1	2.218



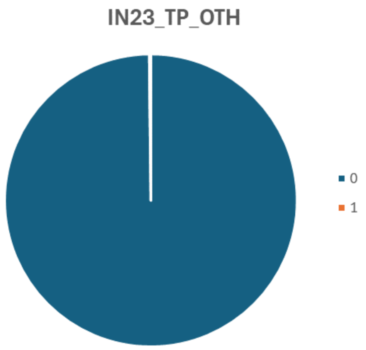
Source: Own elaboration

SMEs details

Only 0,2% of the accounts have other participant or relationship.

Table 184 and Figure 175. Distribution of SMEs accounts by other participant or relationship flag

IN23_TP_OTH	Record_Count
0	195.612
1	350



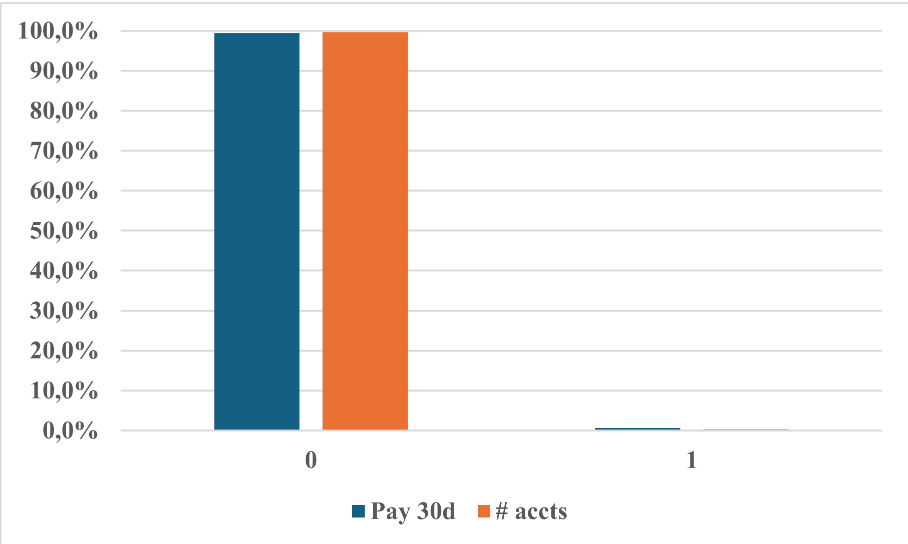
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

In this case, we observe that there is very limited correlation between the other participant or relationship flag and the payment in the following 30 days (the distribution of accounts and payments is similar for the different other participant or relationship flags). The information value (IV) of this variable is 0.00 for Individuals and 0.00 for SMEs.

Table 185 and Figure 176. Bivariant analysis other participant or relationship flag – payment 30 days for Individuals

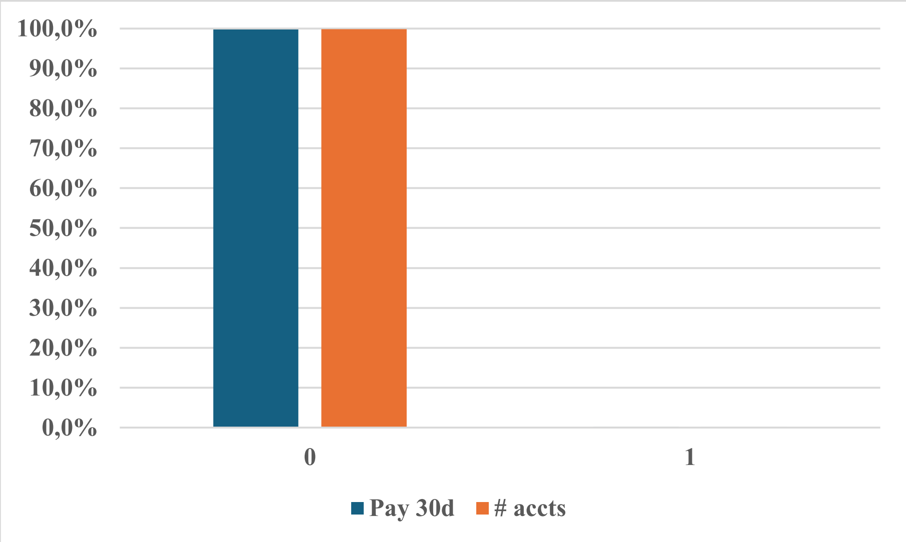
Other participant or relationship flag	Pay 30d	# accts
0	99,4%	99,7%
1	0,6%	0,3%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 186 and Figure 177. Bivariant analysis other participant or relationship flag – payment 30 days for SMEs

Other participant or relationship flag	Pay 30d	# accts
0	99,8%	99,8%
1	0,2%	0,2%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.19.- Variable "IN24_NumPersonas" – Number of participants

Description

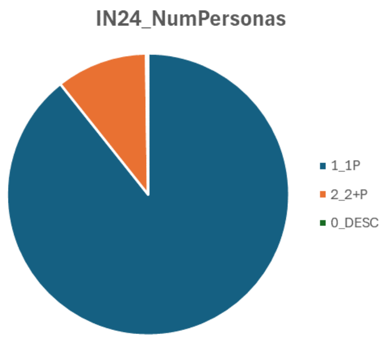
Number of participants linked to each account. This variable was included in the data set, as normally shows correlation with the probability to collect. The number of participants has a positive correlation with the probability to collect: with a higher number of participants, the probability to collect increases. Finally, this variable was relevant in our modelling only for the Individuals Pre Covid-19 model.

Individuals' details

89% of the accounts have just one participant, 11% has two or more participants and just 0,2% of the accounts is unknown or have no information.

Table 187 and Figure 178. Distribution of Individuals accounts by number of participants

IN24_NumPersonas	Record_Count
1_1P	603.861
2_2+P	70.691
0_DESC	1.476



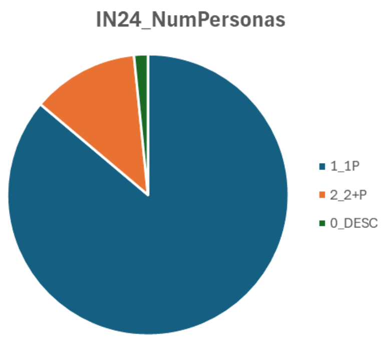
Source: Own elaboration

SMEs details

86% of the accounts have just one participant, 12% has two or more participants and 2% of the accounts is unknown or have no information.

Table 188 and Figure 179. Distribution of SMEs accounts by number of participants

IN24_NumPersonas	Record_Count
1_1P	168.834
2_2+P	24.004
0_DESC	3.124



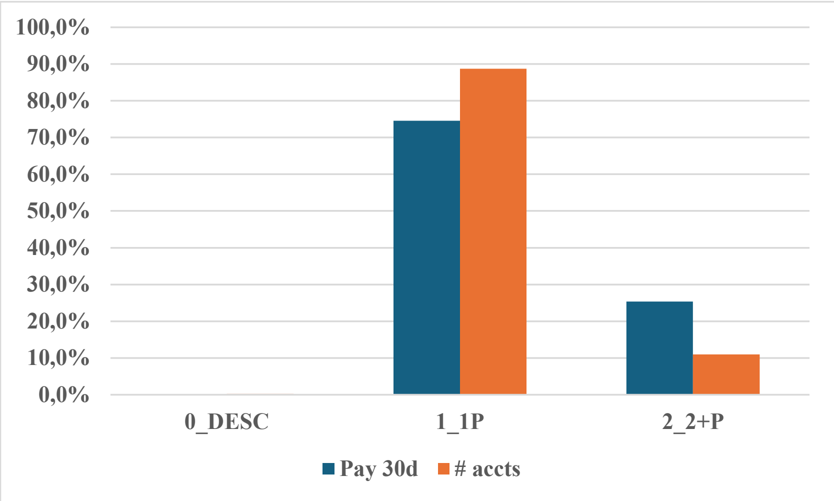
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

In this case, we observe that there is correlation between the number of participants and the payment in the following 30 days (the distribution of accounts and payments is not similar for the number of participants). For both Individuals and SMEs, the probability for a payment is higher when there are 2 or more participants. The information value (IV) of this variable is 0.15 for Individuals and 0.27 for SMEs.

Table 189 and Figure 180. Bivariant analysis number of participants – payment 30 days for Individuals

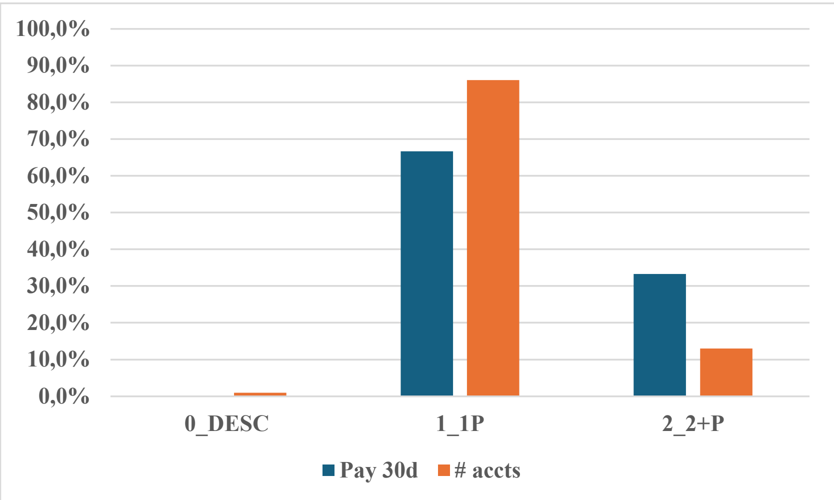
# of participants	Pay 30d	# accts
0_DESC	0,0%	0,2%
1_1P	74,6%	88,7%
2_2+P	25,4%	11,0%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 190 and Figure 181. Bivariant analysis number of participants – payment 30 days for SMEs

# of participants	Pay 30d	# accts
0_DESC	0,0%	1,0%
1_1P	66,6%	86,0%
2_2+P	33,3%	13,0%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.20.- Variable "IN25_TotalTelefonos" – Number of phone numbers

Description

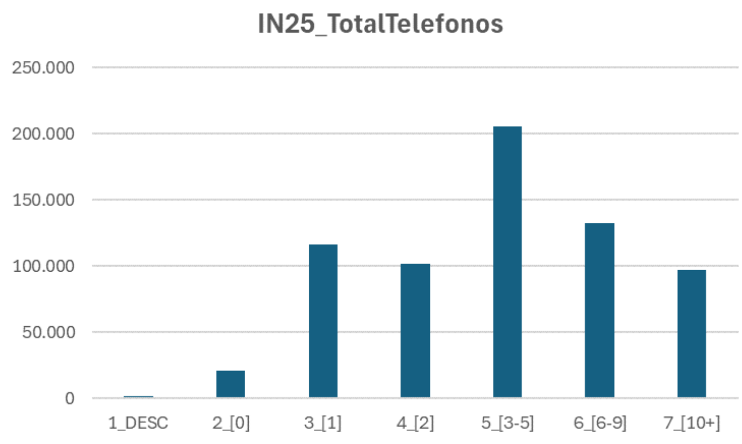
This variable contains the total number of phone numbers available for each account, grouped in several ranges; and it was included in the data set, as normally shows correlation with the probability to collect. The number of phone numbers has a positive correlation with the probability to collect: with a higher number of phone numbers, the probability to collect increases. Finally, this variable was relevant in our modelling for several of the models developed: Individuals Global, Individuals Pre Covid-19, Individuals Post Covid-19, SMEs Global, SMEs Pre Covid-19 and SMEs Post Covid-19.

Individuals' details

64% of the accounts have three or more phone numbers available, and 14% has ten or more phone numbers. 3% of the accounts has no phone number available and only 0.2% of the accounts with unknown or null value in this field.

Table 191 and Figure 182. Distribution of Individuals accounts by number of phone numbers

IN25_TotalTelefonos	Record_Count
1_DESC	1.476
2_[0]	21.057
3_[1]	116.504
4_[2]	101.306
5_[3-5]	205.556
6_[6-9]	132.786
7_[10+]	97.343



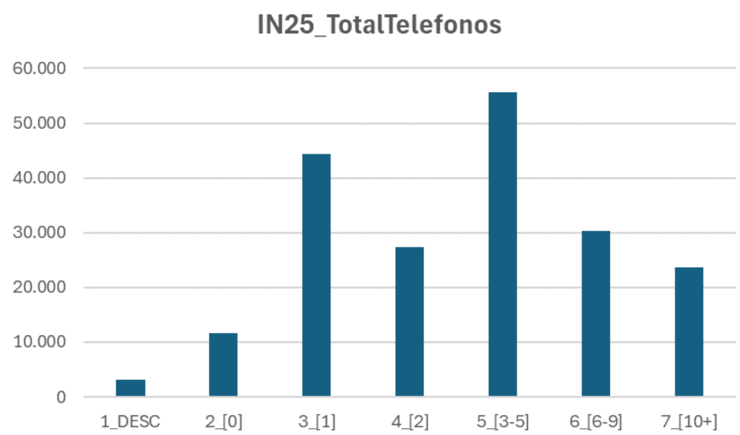
Source: Own elaboration

SMEs details

64% of the accounts have three or more phone numbers available, and 14% has ten or more phone numbers. 3% of the accounts has no phone number available and only 0.2% of the accounts with unknown or null value in this field.

Table 192 and Figure 183. Distribution of SMEs accounts by number of phone numbers

IN25_TotalTelefonos	Record_Count
1_DESC	3.124
2_[0]	11.604
3_[1]	44.334
4_[2]	27.360
5_[3-5]	55.590
6_[6-9]	30.259
7_[10+]	23.691



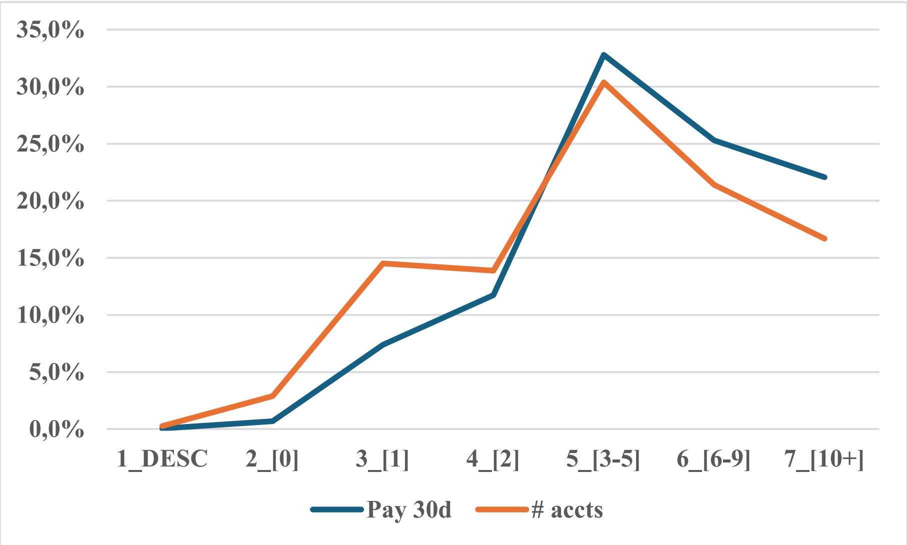
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

In this case, we observe that there is very limited correlation between the number of phone numbers and the payment in the following 30 days (the distribution of accounts and payments is similar for the number of phone numbers). The probability for a payment is higher when the number of phone numbers is 3 or higher (in both verticals, Individuals and SMEs). The information value (IV) of this variable is 0.11 for Individuals and 0.30 for SMEs.

Table 193 and Figure 184. Bivariant analysis number of phone numbers – payment 30 days for Individuals

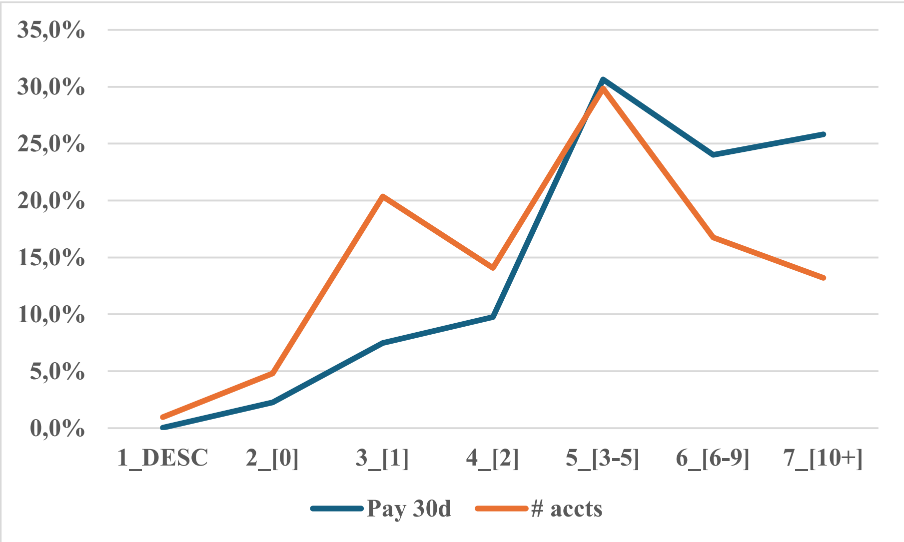
# of phone numbers	Pay 30d	# accts
1_DESC	0,0%	0,2%
2_[0]	0,7%	2,9%
3_[1]	7,4%	14,5%
4_[2]	11,7%	13,9%
5_[3-5]	32,8%	30,4%
6_[6-9]	25,3%	21,4%
7_[10+]	22,1%	16,7%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 194 and Figure 185. Bivariant analysis number of phone numbers – payment 30 days for SMEs

# of phone numbers	Pay 30d	# accts
1_DESC	0,0%	1,0%
2_[0]	2,3%	4,8%
3_[1]	7,5%	20,4%
4_[2]	9,8%	14,1%
5_[3-5]	30,6%	29,8%
6_[6-9]	24,0%	16,7%
7_[10+]	25,8%	13,2%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.21.- Variable "IN26_CCAA" – Region

Description

This variable contains the information about the Region ("Comunidades Autónomas") of each account, by grouping the provinces information contained in the variable "IN20_Provincia".

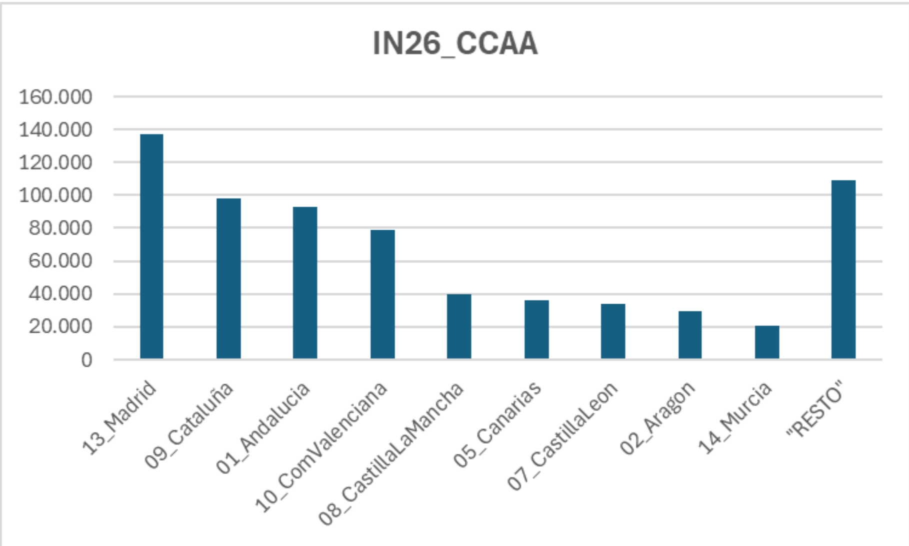
This variable was included in the data set, as normally shows correlation with the probability to collect. The region where the debt was originated have a clear correlation with the probability to collect, as the demographic and macroeconomics factors are very different among the regions, and these factors directly influence in the probability to collect. Certain cities, provinces and regions have a higher probability to collect than others. In our research, this variable was relevant for the Individuals Global, Individuals Pre Covid-19, Individuals Covid-19, SMEs Pre Covid-19 and SMEs Covid-19 models.

Individuals' details

Madrid and Catalunya sum up for 35% of the total accounts, and Andalucia and Valencia counts for 25% of the total account among the two of them. The rest of the accounts is well spread among the rest of the Regions.

Table 195 and Figure 186. Distribution of Individuals accounts by region

IN26_CCAA	Record_Count
13_Madrid	137.147
09_Cataluña	97.894
01_Andalucia	92.557
10_ComValenciana	78.973
08_CastillaLaMancha	39.888
05_Canarias	35.850
07_CastillaLeon	34.059
02_Aragon	29.540
14_Murcia	20.904



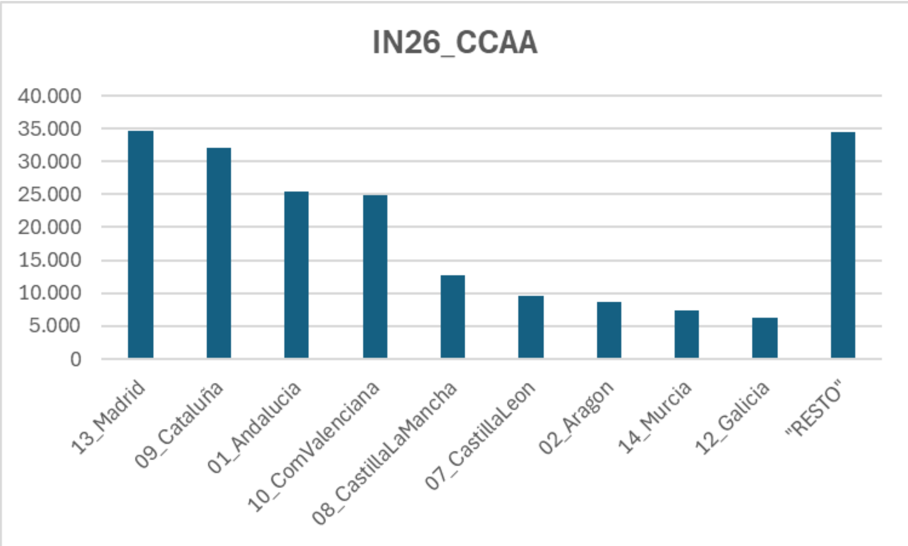
Source: Own elaboration

SMEs details

Madrid and Catalunya sum up for 34% of the total accounts, and Andalucia and Valencia counts for 26% of the total account among the two of them. The rest of the accounts is well spread among the rest of the Regions.

Table 196 and Figure 187. Distribution of SMEs accounts by region

IN26_CCAA	Record_Count
13_Madrid	34.654
09_Cataluña	32.028
01_Andalucia	25.509
10_ComValenciana	24.807
08_CastillaLaMancha	12.631
07_CastillaLeon	9.663
02_Aragon	8.659
14_Murcia	7.408
12_Galicia	6.198
05_Canarias	6.008



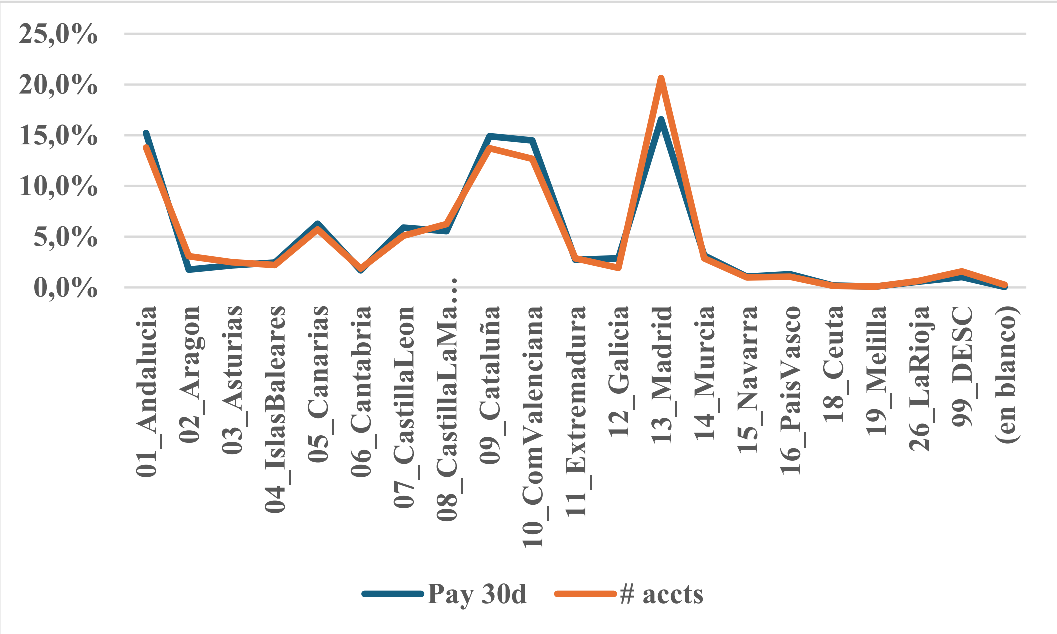
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

In this case, we observe that there is certain correlation between the region and the payment in the following 30 days (the distribution of accounts and payments is not similar for the different regions). The probability for a payment varies widely with the region. The information value (IV) of this variable is 0.04 for Individuals and 0.08 for SMEs.

Table 197 and Figure 188. Bivariant analysis region – payment 30 days for Individuals

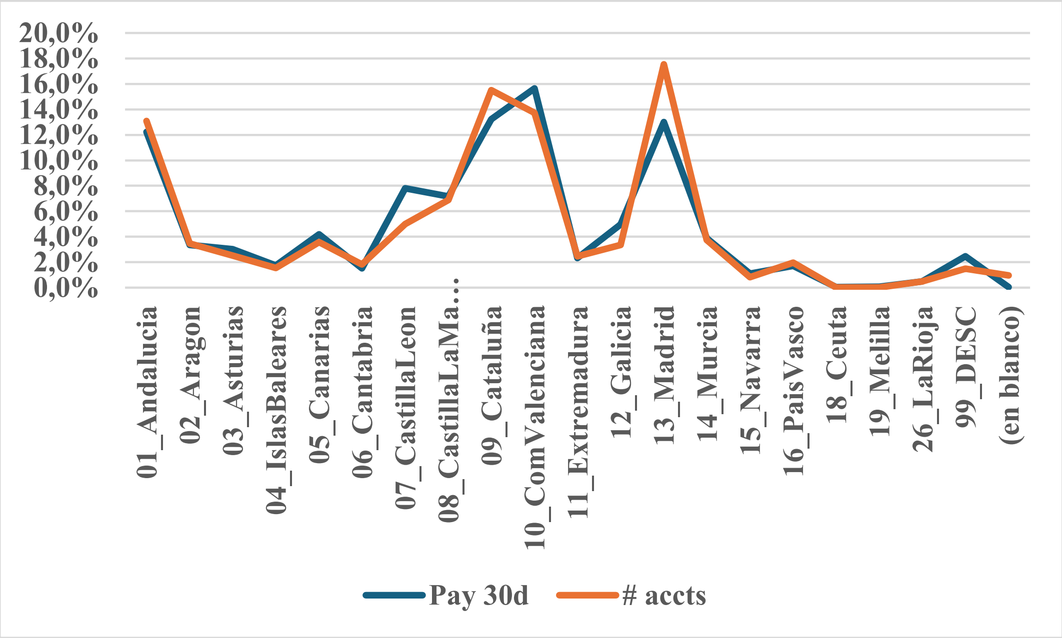
Region	Pay 30d	# accts
01_Andalucia	15,2%	13,8%
02_Aragon	1,7%	3,1%
03_Asturias	2,2%	2,5%
04_IslasBaleares	2,5%	2,2%
05_Canarias	6,3%	5,7%
06_Cantabria	1,7%	1,9%
07_CastillaLeon	5,9%	5,1%
08_CastillaLaMancha	5,6%	6,2%
09_Cataluña	14,9%	13,7%
10_ComValenciana	14,5%	12,7%
11_Extremadura	2,7%	2,9%
12_Galicia	2,9%	1,9%
13_Madrid	16,6%	20,6%
14_Murcia	3,2%	2,9%
15_Navarra	1,1%	1,0%
16_PaisVasco	1,3%	1,1%
18_Ceuta	0,2%	0,2%
19_Melilla	0,1%	0,1%
26_LaRioja	0,6%	0,6%
99_DESC	1,0%	1,6%
(en blanco)	0,0%	0,2%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 198 and Figure 189. Bivariant analysis region – payment 30 days for SMEs

Region	Pay 30d	# accts
01_Andalucia	12,2%	13,1%
02_Aragon	3,3%	3,5%
03_Asturias	3,0%	2,5%
04_IslasBaleares	1,7%	1,5%
05_Canarias	4,2%	3,6%
06_Cantabria	1,5%	1,8%
07_CastillaLeon	7,8%	5,0%
08_CastillaLaMancha	7,2%	6,9%
09_Cataluña	13,2%	15,5%
10_ComValenciana	15,6%	13,7%
11_Extremadura	2,3%	2,5%
12_Galicia	5,0%	3,3%
13_Madrid	13,0%	17,5%
14_Murcia	3,9%	3,7%
15_Navarra	1,1%	0,8%
16_PaisVasco	1,7%	2,0%
18_Ceuta	0,0%	0,0%
19_Melilla	0,1%	0,0%
26_LaRioja	0,5%	0,5%
99_DESC	2,5%	1,5%
(en blanco)	0,0%	1,0%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.22.- Variable “IN29_SMS_L30D” – SMS sent in the last 30 days flag

Description

This variable reflects if the account has had an SMS sent in the last 30 days for every month of the analysis period. For the construction of this flag, it was used the last available data point for any given month.

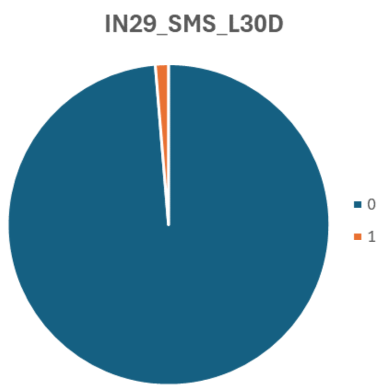
This variable was included in the data set, as normally shows correlation with the probability to collect. The collections and recoveries actions (SMS, calls…) have a positive correlation with the probability to collect: with a higher number of collection actions, the probability to collect increases. As explained in Section 4.4.3 of this document, this variable was eliminated for our modelling due to the excessive correlation with the variable to predict.

Individuals’ details

Analysing the last available data point for each bionomy month-account, we observe that just 1.3% of the accounts have an SMS sent in the last 30 days.

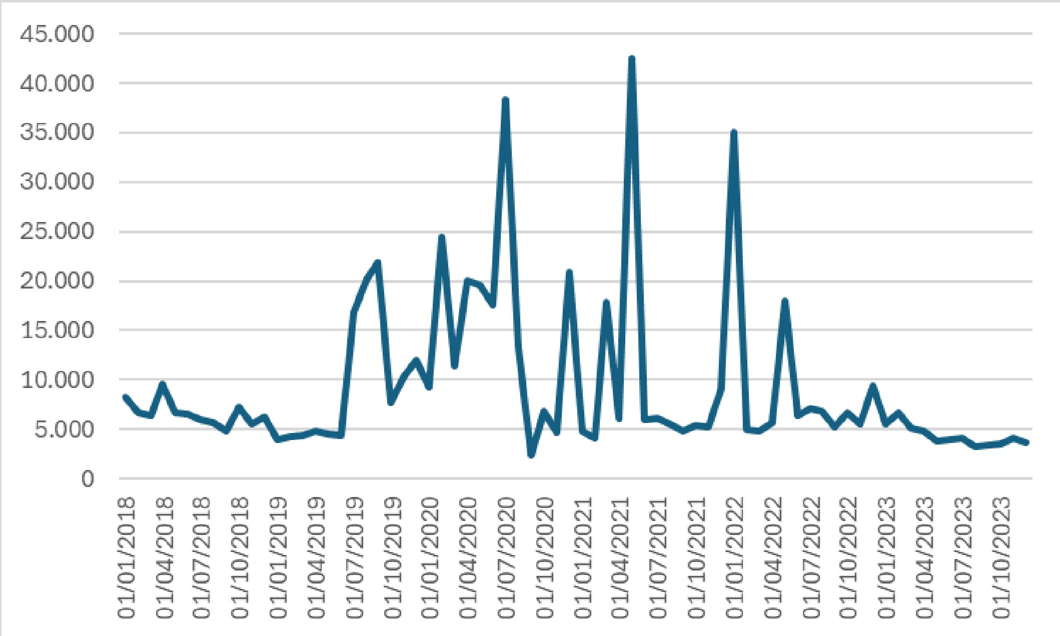
Table 199 and Figure 190. Distribution of Individuals accounts by SMS L30D flag

IN29_SMS_L30D	Record_Count
0	667.127
1	8.901



Source: Own elaboration

Figure 191. Monthly evolution of Individuals accounts by SMS L30D flag



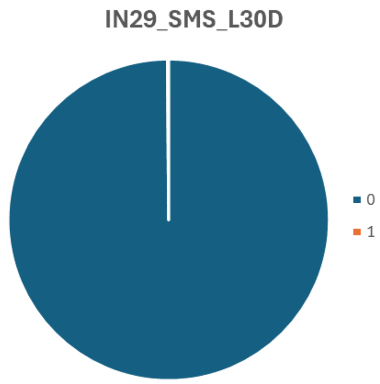
Source: Own elaboration

SMEs details

Analysing the last available data point for each bionomy month-account, we observe that just 0.1% of the accounts have an SMS sent in the last 30 days.

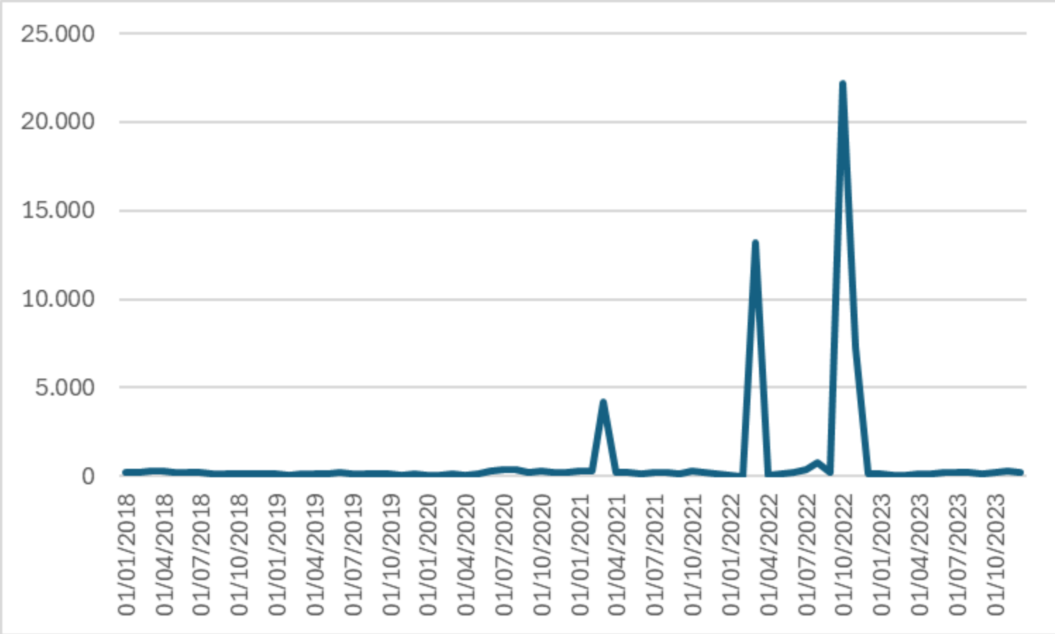
Table 200 and Figure 192. Distribution of SMEs accounts by SMS L30D flag

IN29_SMS_L30D	Record_Count
0	195.722
1	240



Source: Own elaboration

Figure 193. Monthly evolution of SMEs accounts by SMS L30D flag



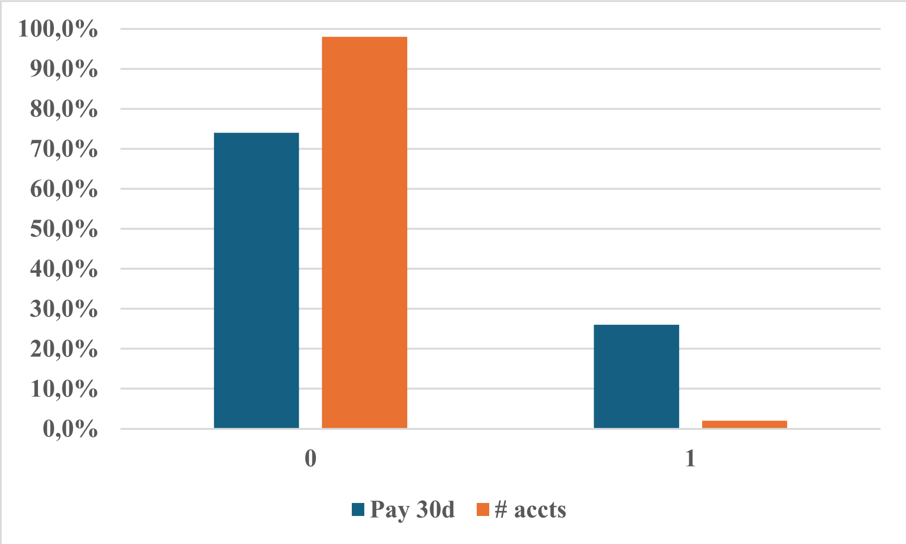
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

For this variable, we observe that there is a clear correlation between the SMS sent in the last 30 days flag and the payment in the following 30 days (the distribution of accounts and payments is not similar for the positive and negative results of the flag). Both in Individuals and SMEs, the probability for a payment is much higher when the flag result is positive. The information value (IV) of this variable is 0.68 for Individuals and 0.05 for SMEs.

Table 201 and Figure 194. Bivariant analysis SMS sent last 30d flag – payment 30 days for Individuals

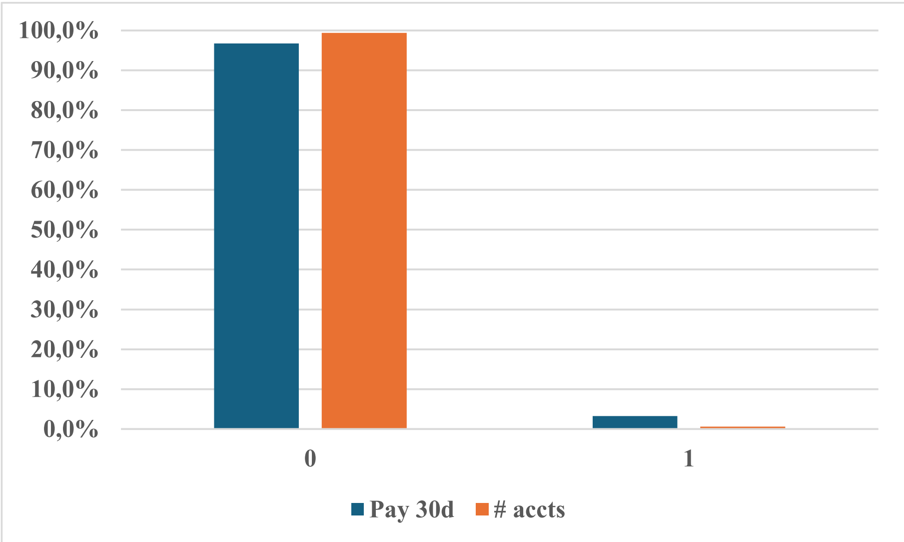
SMS sent last 30d flag	Pay 30d	# accts
0	74,0%	98,0%
1	26,0%	2,0%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 202 and Figure 195. Bivariant analysis SMS sent last 30d flag – payment 30 days for SMEs

SMS sent last 30d flag	Pay 30d	# accts
0	96,7%	99,4%
1	3,3%	0,6%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.23.- Variable "IN30_N_LLAMADAS_L30D" – Calls in the last 30 days flag

Description

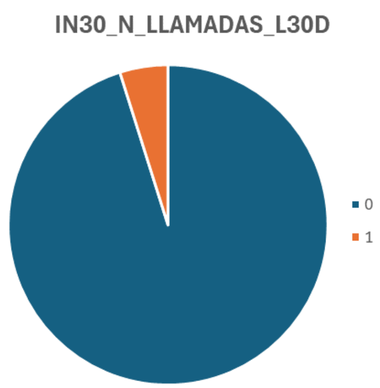
This variable reflects if the account has had a Call in the last 30 days for every month of the analysis period. For the construction of this flag, it was used the last available data point for any given month. And this variable was included in the data set, as normally shows correlation with the probability to collect. The collections and recoveries actions (SMS, calls…) have a positive correlation with the probability to collect: with a higher number of collection actions, the probability to collect increases. This variable was relevant in our modelling only for the Individuals Post Covid-19 and SMEs Post Covid-19 models.

Individuals' details

Analysing the last available data point for each bionomy month-account, we observe that just 4.9% of the accounts had a call in the last 30 days.

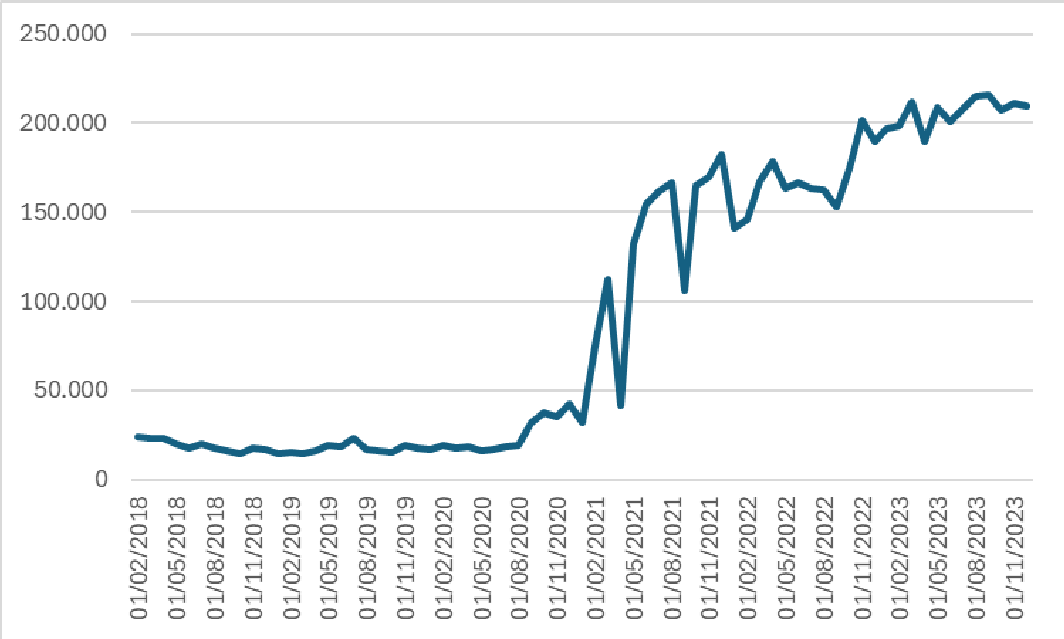
Table 203 and Figure 196. Distribution of Individuals accounts by Call L30D flag

IN30_N_LLAMADAS_L30D	Record_Count
0	643.095
1	32.933



Source: Own elaboration

Figure 197. Monthly evolution of Individuals accounts by Call L30D flag



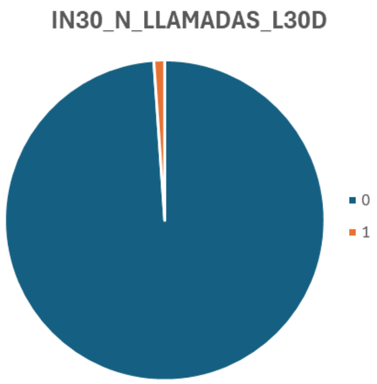
Source: Own elaboration

SMEs details

Analysing the last available data point for each bionomy month-account, we observe that just 1.1% of the accounts had a call in the last 30 days.

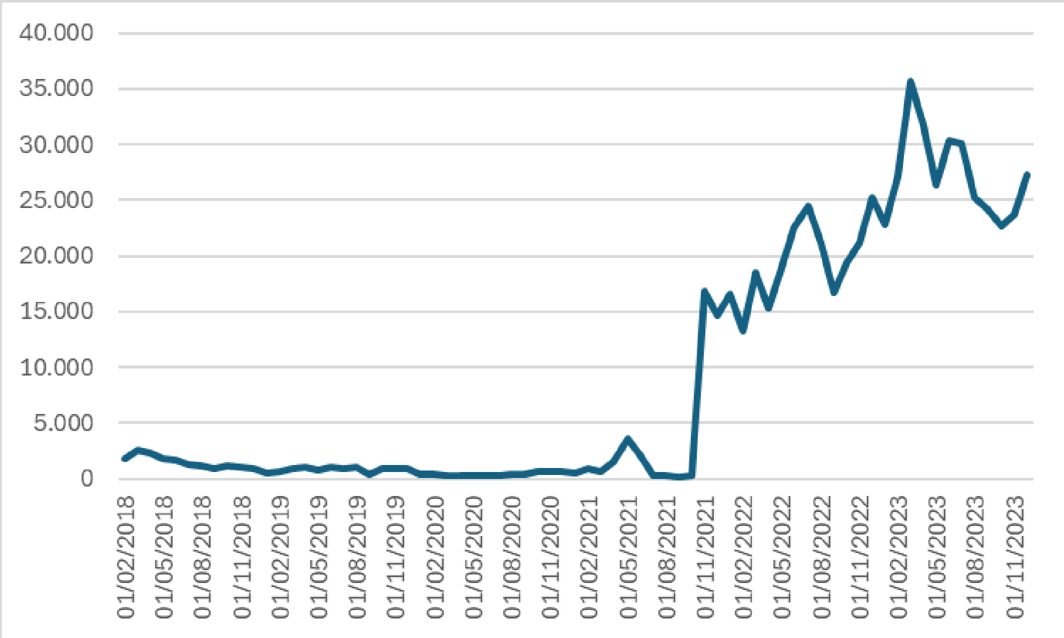
Table 204 and Figure 198. Distribution of SMEs accounts by Call L30D flag

IN30_N_LLAMADAS_L30D	Record_Count
0	193.814
1	2.148



Source: Own elaboration

Figure 199. Monthly evolution of SMEs accounts by Call L30D flag



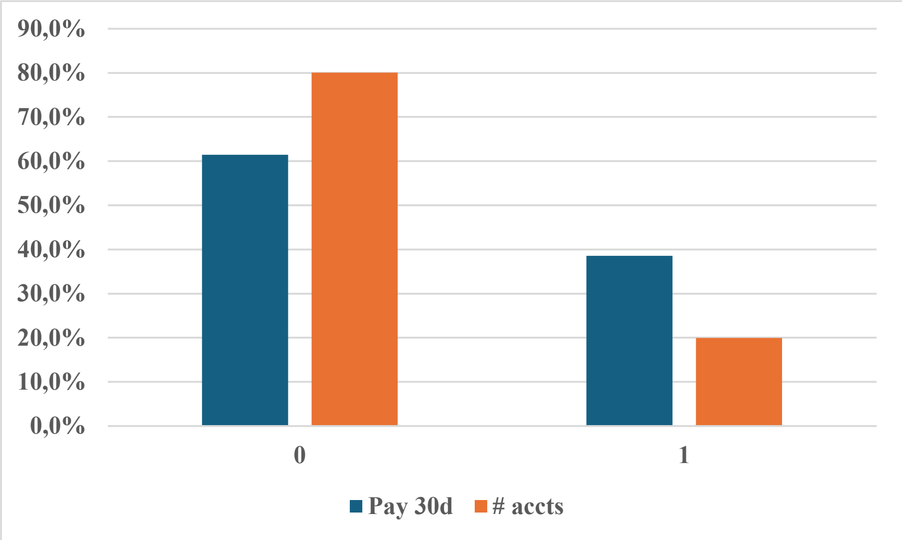
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

For this variable, we observe that there is correlation between the calls in the last 30 days flag and the payment in the following 30 days (the distribution of accounts and payments is not similar for the positive and negative results of the flag). Both in Individuals and SMEs, the probability for a payment is higher when the flag result is positive. The information value (IV) of this variable is 0.17 for Individuals and 0.07 for SMEs.

Table 205 and Figure 200. Bivariant analysis calls last 30d flag – payment 30 days for Individuals

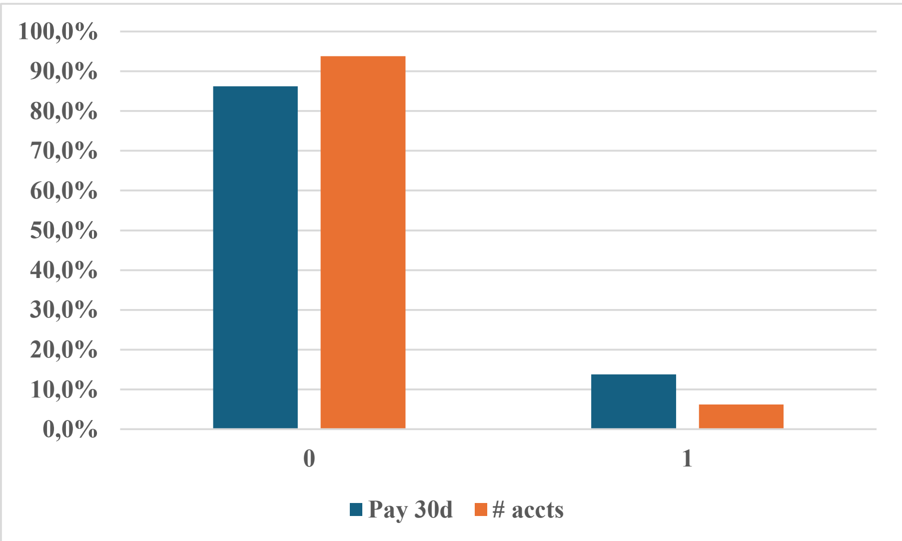
Calls last 30d flag	Pay 30d	# accts
0	61,5%	80,0%
1	38,5%	20,0%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 206 and Figure 201. Bivariant analysis calls last 30d flag – payment 30 days for SMEs

Calls last 30d flag	Pay 30d	# accts
0	86,2%	93,8%
1	13,8%	6,2%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.24.- Variable “IN31_RPC_L30D” – Right-party contact in the last 30 days flag

Description

This variable reflects if the account has had a Right-party Contact (RPC) in the last 30 days for every month of the analysis period. For the construction of this flag, it was used the last available data point for any given month.

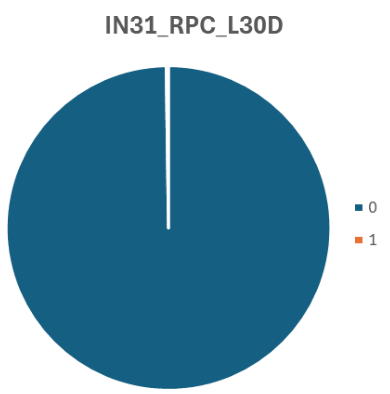
This variable was included in the data set, as normally shows correlation with the probability to collect. The collections and recoveries actions (SMS, calls…) have a positive correlation with the probability to collect: with a higher number of collection actions, the probability to collect increases. Finally, in our modelling this variable was relevant for the models developed.

Individuals’ details

Analysing the last available data point for each bionomy month-account, we observe that only 0.2% of the accounts had a right-party contact in the last 30 days.

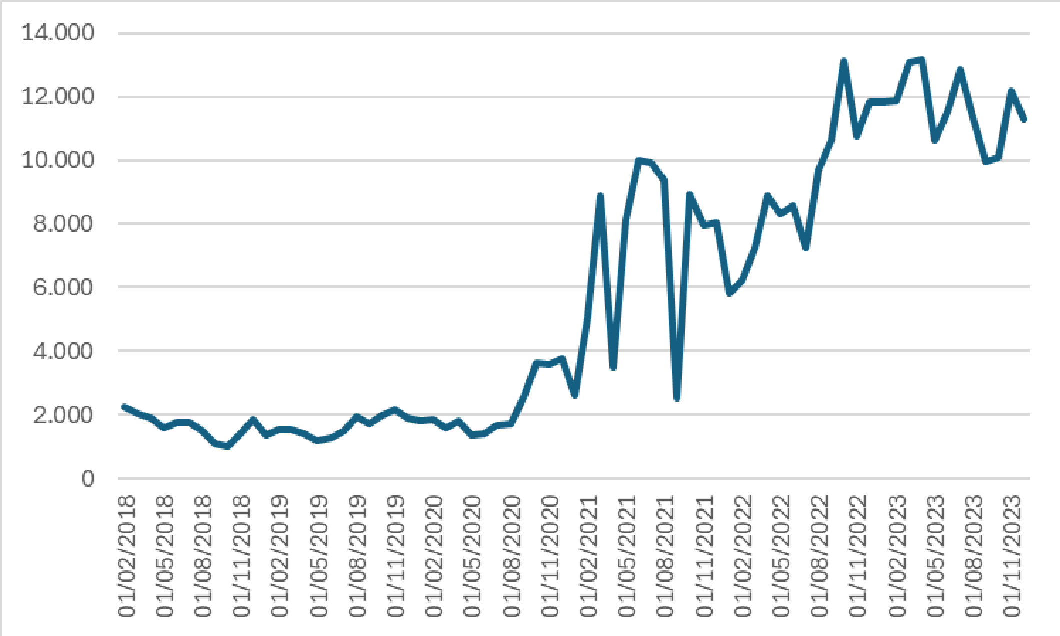
Table 207 and Figure 202. Distribution of Individuals accounts by RPC L30D flag

IN31_RPC_L30D	Record_Count
0	674.503
1	1.525



Source: Own elaboration

Figure 203. Monthly evolution of Individuals accounts by RPC L30D flag



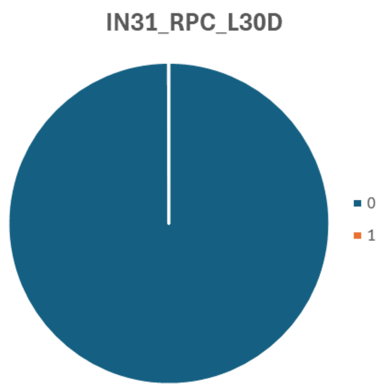
Source: Own elaboration

SMEs details

Analysing the last available data point for each bionomy month-account, we observe that only 0.04% of the accounts had a right-party contact in the last 30 days.

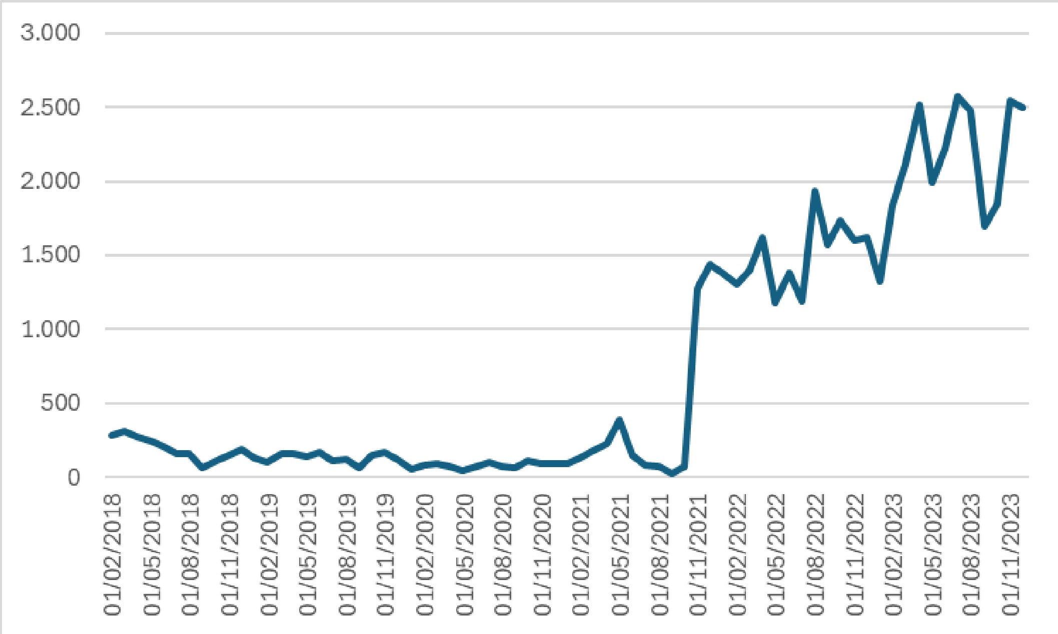
Table 208 and Figure 204. Distribution of SMEs accounts by RPC L30D flag

IN31_RPC_L30D	Record_Count
0	195.887
1	75



Source: Own elaboration

Figure 205. Monthly evolution of SMEs accounts by RPC L30D flag



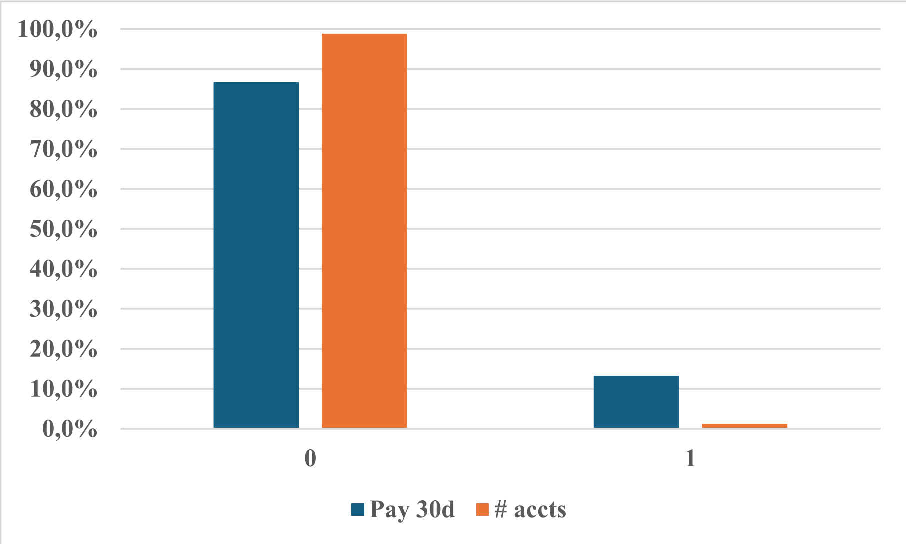
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

In this case, we observe that there is correlation between the RPC in the last 30 days flag and the payment in the following 30 days (the distribution of accounts and payments is not similar for the positive and negative results of the flag). Both in Individuals and SMEs, the probability for a payment is higher when the flag result is positive. The information value (IV) of this variable is 0.31 for Individuals and 0.08 for SMEs.

Table 209 and Figure 206. Bivariant analysis RPC last 30d flag – payment 30 days for Individuals

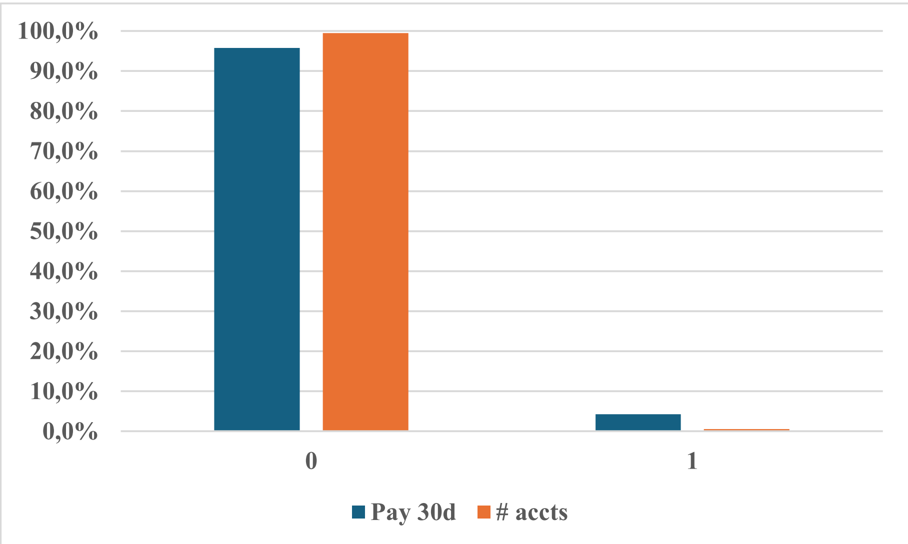
RPC last 30d flag	Pay 30d	# accts
0	86,8%	98,8%
1	13,2%	1,2%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 210 and Figure 207. Bivariant analysis RPC last 30d flag – payment 30 days for SMEs

RPC last 30d flag	Pay 30d	# accts
0	95,8%	99,5%
1	4,2%	0,5%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.25.- Variable "IN32_PTP _L30D" – Promise to pay in the last 30 days flag

Description

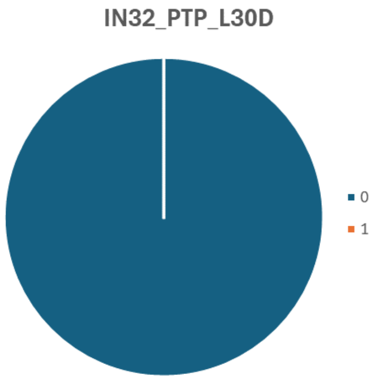
This variable reflects if the account has had a Promise to pay (PTP) in the last 30 days for every month of the analysis period. For the construction of this flag, it was used the last available data point for any given month. And this variable was included in the data set, as normally shows correlation with the probability to collect. The collections and recoveries actions (SMS, calls…) have a positive correlation with the probability to collect: with a higher number of collection actions, the probability to collect increases. This variable was not relevant neither for the Individuals models, nor the SMEs ones.

Individuals' details

Analysing the last available data point for each bionomy month-account, we observe that only 20 of the total accounts had a promise to pay in the last 30 days.

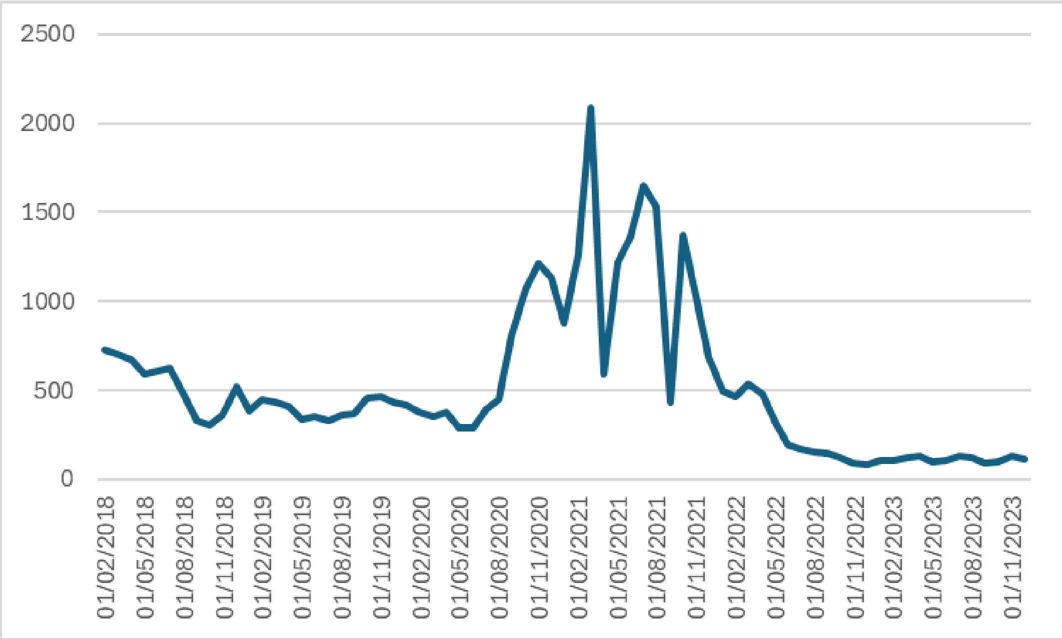
Table 211 and Figure 208. Distribution of Individuals accounts by PTP L30D flag

IN32_PTP_L30D	Record_Count
0	676.008
1	20



Source: Own elaboration

Figure 209. Monthly evolution of Individuals accounts by PTP L30D flag



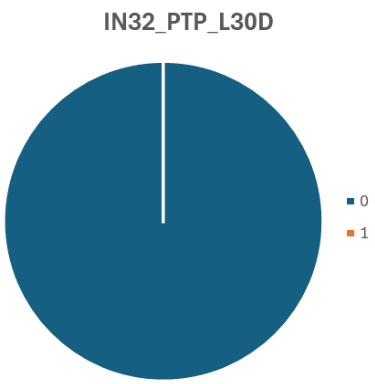
Source: Own elaboration

SMEs details

Analysing the last available data point for each bionomy month-account, we observe that none of the accounts had a promise to pay in the last 30 days.

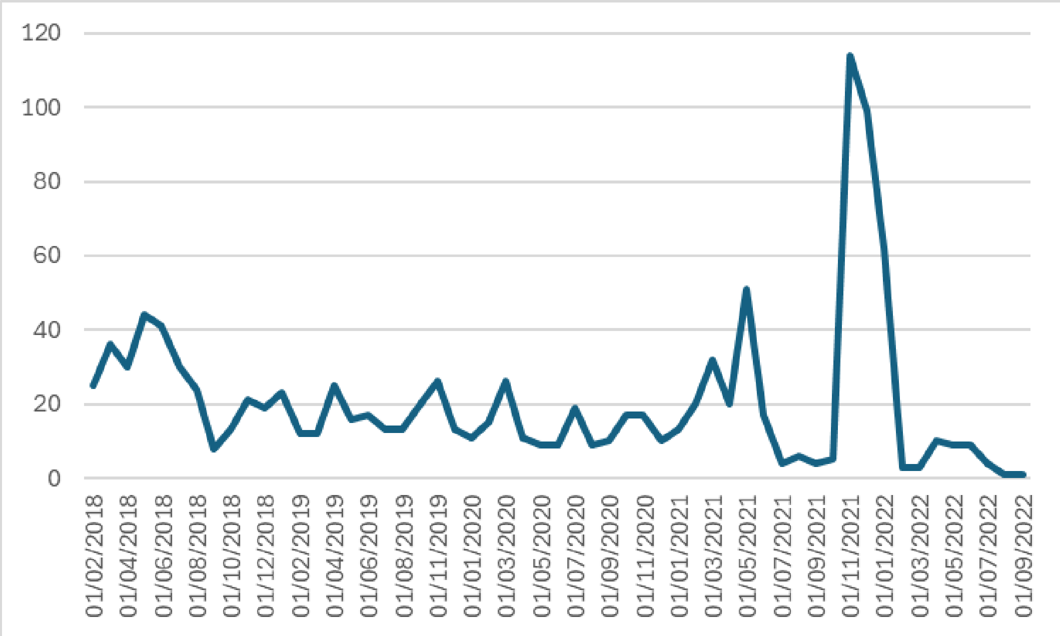
Table 212 and Figure 210. Distribution of SMEs accounts by PTP L30D flag

IN32_PTP_L30D	Record_Count
0	195.962
1	0



Source: Own elaboration

Figure 211. Monthly evolution of SMEs accounts by PTP L30D flag



* No new PTP after sep-22

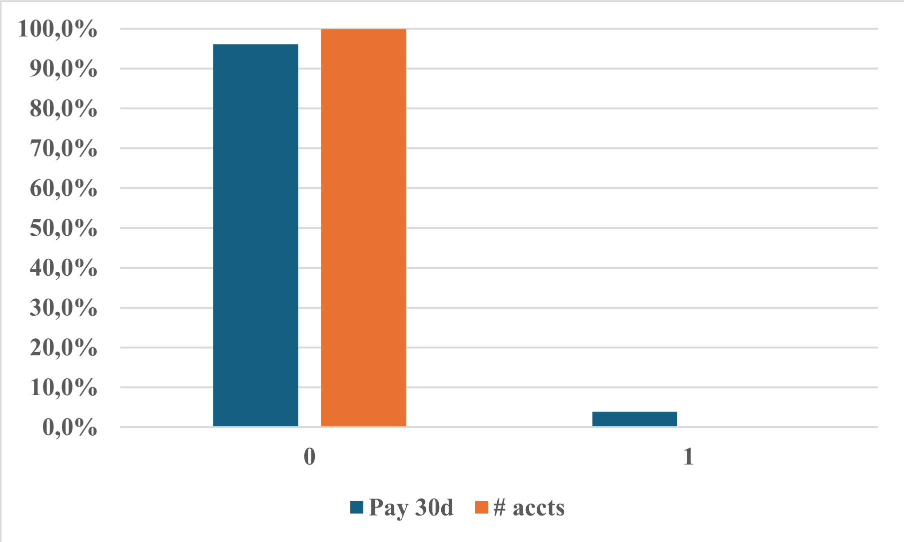
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

For this variable, we observe that there is correlation between the PTP in the last 30 days flag and the payment in the following 30 days (the distribution of accounts and payments is not similar for the positive and negative results of the flag). Both in Individuals and SMEs, the probability for a payment is higher when the flag result is positive. The information value (IV) of this variable is 0.14 for Individuals and 0.01 for SMEs.

Table 213 and Figure 212. Bivariant analysis PTP last 30d flag – payment 30 days for Individuals

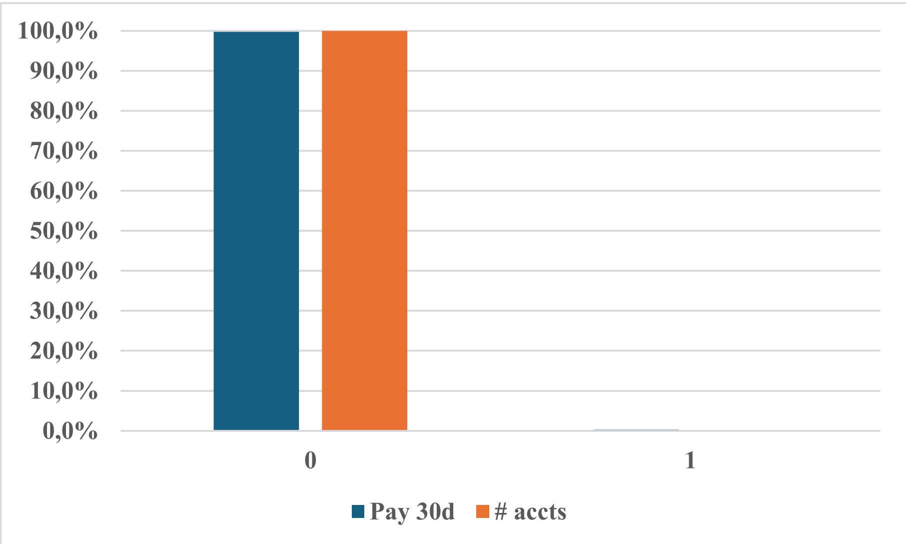
PTP last 30d flag	Pay 30d	# accts
0	96,1%	99,9%
1	3,9%	0,1%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 214 and Figure 213. Bivariant analysis PTP last 30d flag – payment 30 days for SMEs

PTP last 30d flag	Pay 30d	# accts
0	99,7%	100,0%
1	0,3%	0,0%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.26.- Variable "IN34_N_LLAMADAS_L90D" – Calls in the last 90 days flag

Description

This variable reflects if the account has had a Call in the last 90 days for every month of the analysis period. For the construction of this flag, it was used the last available data point for any given month.

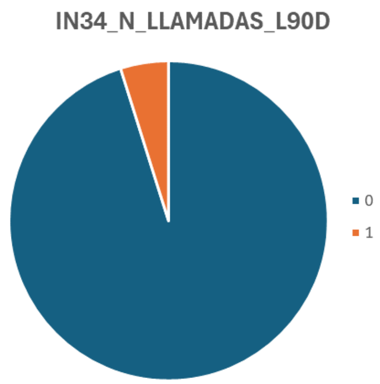
This variable was included in the data set, as normally shows correlation with the probability to collect. The collections and recoveries actions (SMS, calls…) have a positive correlation with the probability to collect: with a higher number of collection actions, the probability to collect increases. Within the models with in this research, this variable was relevant for the Individuals Global, Individuals Covid-19, Individuals Post Covid-19 and SMEs Post Covid-19 models.

Individuals' details

Analysing the last available data point for each bionomy month-account, we observe that just 4.9% of the accounts had a call in the last 90 days.

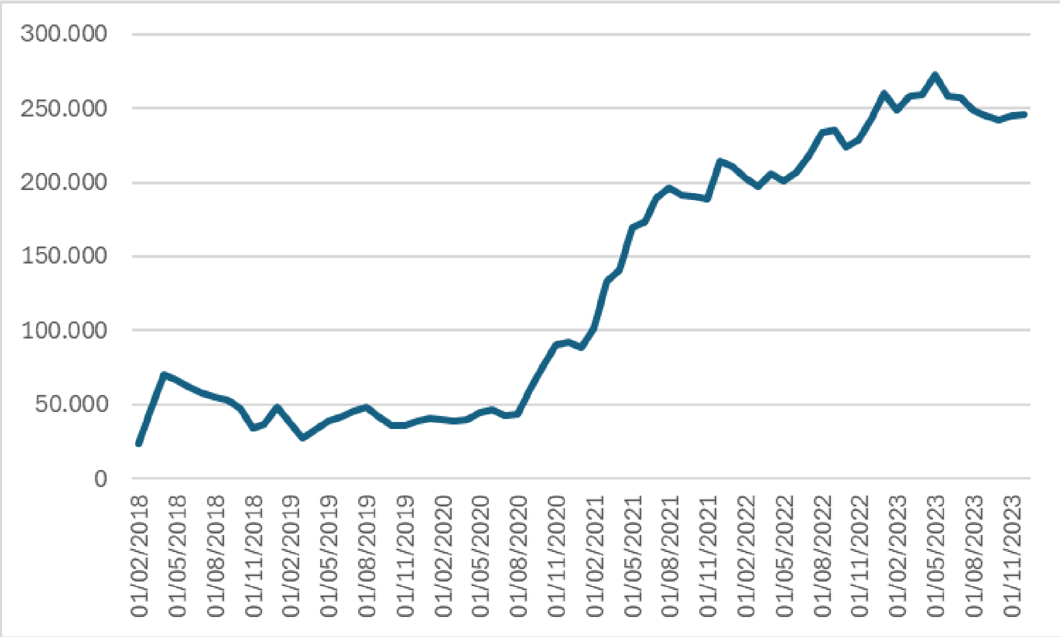
Table 215 and Figure 214. Distribution of Individuals accounts by Call L90D flag

IN34_N_LLAMADAS_L90D	Record_Count
0	643.095
1	32.933



Source: Own elaboration

Figure 215. Monthly evolution of Individuals accounts by Call L90D flag



Source: Own elaboration

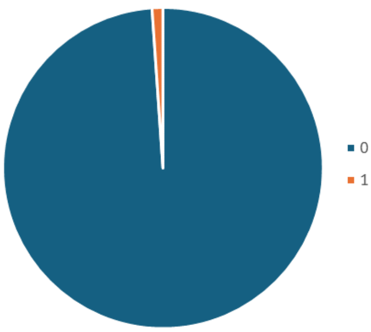
SMEs details

Analysing the last available data point for each bionomy month-account, we observe that just 1.1% of the accounts had a call in the last 30 days.

Table 216 and Figure 216. Distribution of SMEs accounts by Call L30D flag

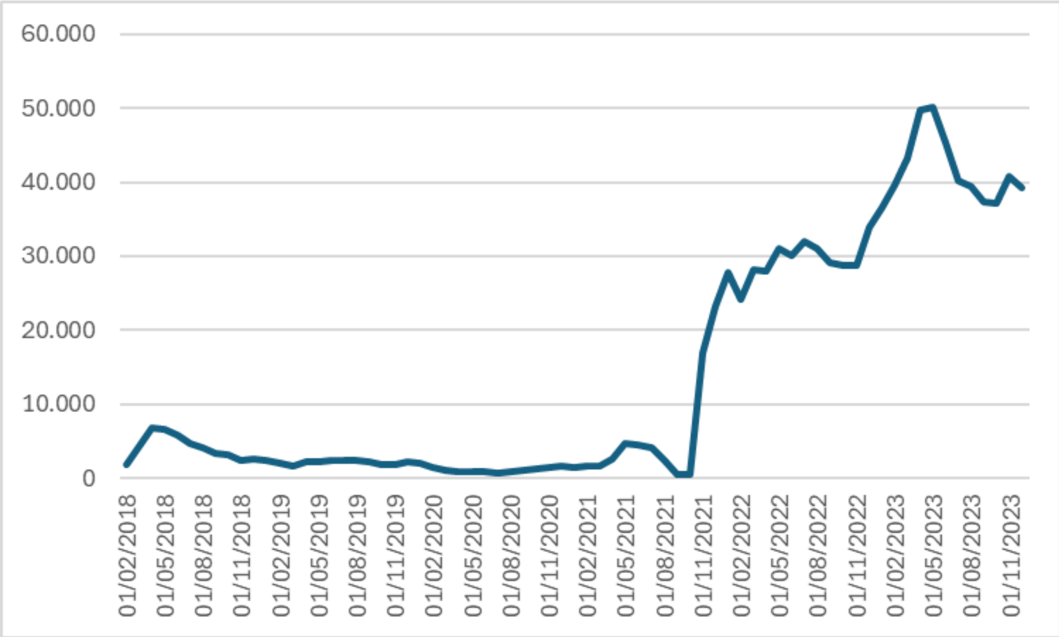
IN34_N_LLAMADAS_L90D	Record_Count
0	193.814
1	2.148

IN34_N_LLAMADAS_L90D



Source: Own elaboration

Figure 217. Monthly evolution of SMEs accounts by Call L30D flag



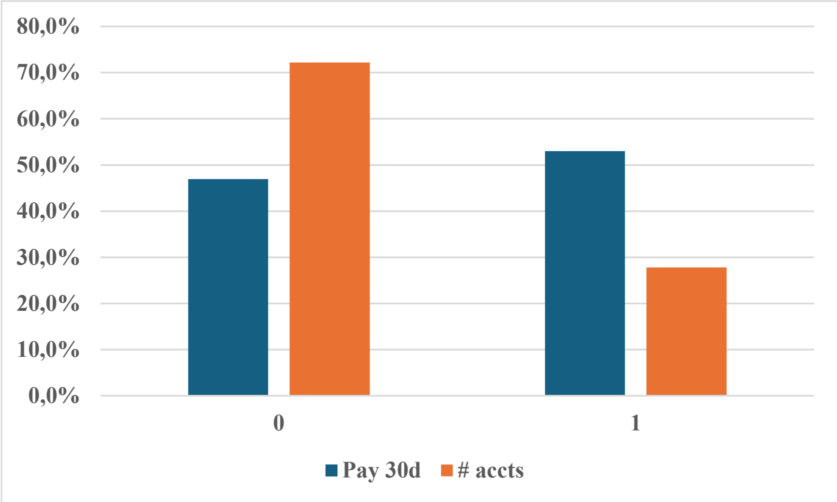
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

For this variable, we observe that there is correlation between the calls in the last 90 days flag and the payment in the following 30 days (the distribution of accounts and payments is not similar for the positive and negative results of the flag). Both in Individuals and SMEs, the probability for a payment is higher when the flag result is positive. The information value (IV) of this variable is 0.27 for Individuals and 0.12 for SMEs.

Table 217 and Figure 218. Bivariant analysis calls last 90d flag – payment 30 days for Individuals

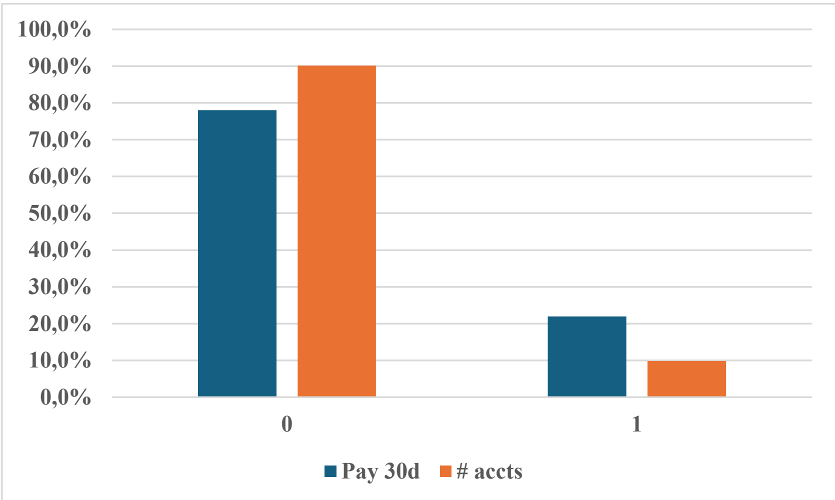
Calls last 90d flag	Pay 30d	# accts
0	47,0%	72,2%
1	53,0%	27,8%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 218 and Figure 219. Bivariant analysis calls last 90d flag – payment 30 days for SMEs

Calls last 90d flag	Pay 30d	# accts
0	78,0%	90,2%
1	22,0%	9,8%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.27.- Variable "IN35_RPC_L90D" – Right-party contact in the last 90 days flag

Description

This variable reflects if the account has had a Right-party Contact (RPC) in the last 90 days for every month of the analysis period. For the construction of this flag, it was used the last available data point for any given month.

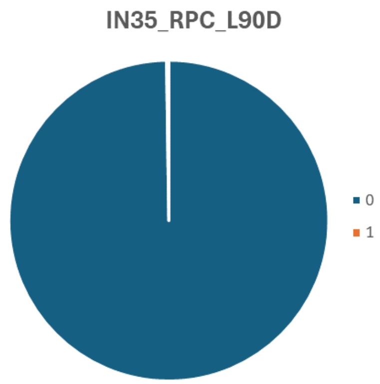
This variable was included in the data set, as normally shows correlation with the probability to collect. The collections and recoveries actions (SMS, calls…) have a positive correlation with the probability to collect: with a higher number of collection actions, the probability to collect increases. This variable was relevant in our modelling only for some Individuals models: Global, Covid-19 and Post Covid-19.

Individuals' details

Analysing the last available data point for each bionomy month-account, we observe that only 0.2% of the accounts had a right-party contact in the last 90 days.

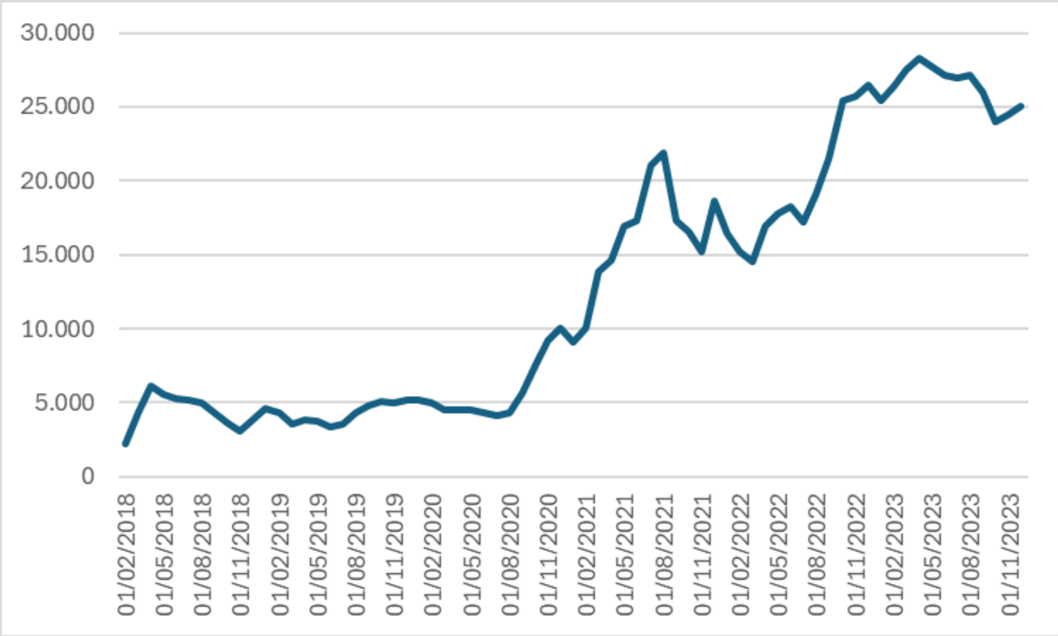
Table 219 and Figure 220. Distribution of Individuals accounts by RPC L90D flag

IN35_RPC_L90D	Record_Count
0	674.503
1	1.525



Source: Own elaboration

Figure 221. Monthly evolution of Individuals accounts by RPC L90D flag



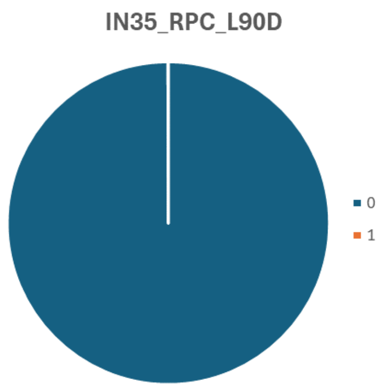
Source: Own elaboration

SMEs details

Analysing the last available data point for each bionomy month-account, we observe that only 0.04% of the accounts had a right-party contact in the last 90 days.

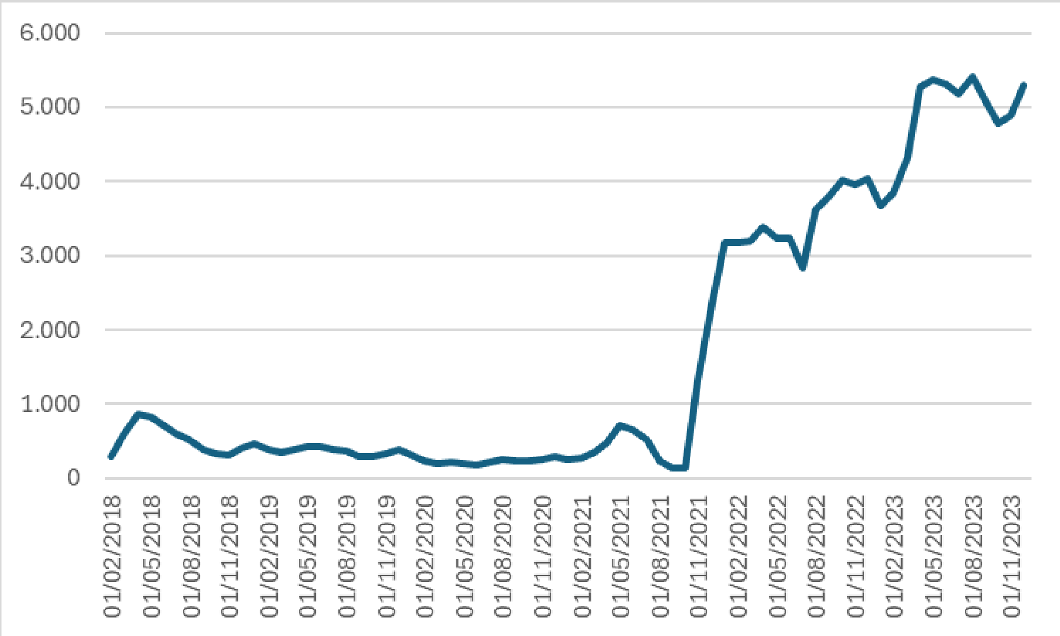
Table 220 and Figure 222. Distribution of SMEs accounts by RPC L90D flag

IN35_RPC_L90D	Record_Count
0	195.887
1	75



Source: Own elaboration

Figure 223. Monthly evolution of SMEs accounts by RPC L90D flag



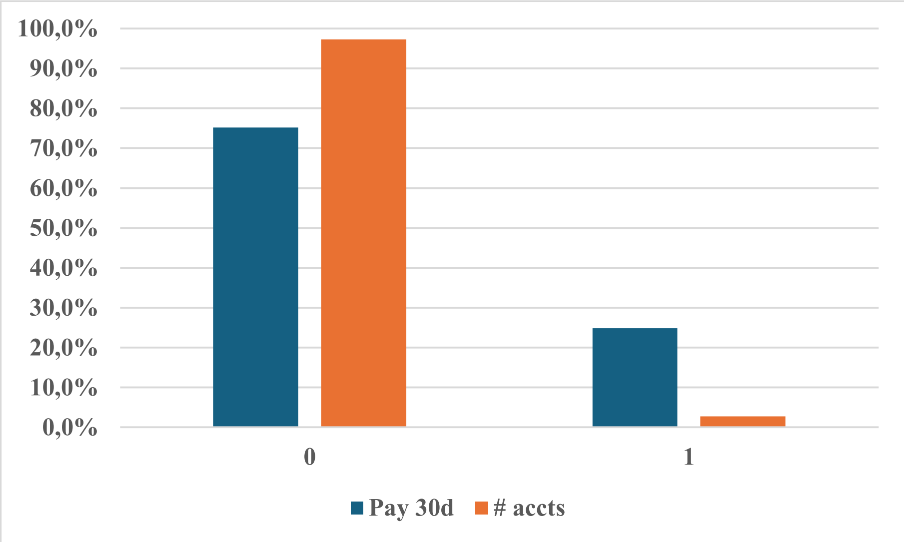
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

In this case, we observe that there is correlation between the RPC in the last 90 days flag and the payment in the following 30 days (the distribution of accounts and payments is not similar for the positive and negative results of the flag). Both in Individuals and SMEs, the probability for a payment is higher when the flag result is positive. The information value (IV) of this variable is 0.55 for Individuals and 0.14 for SMEs.

Table 221 and Figure 224. Bivariant analysis RPC last 90d flag – payment 30 days for Individuals

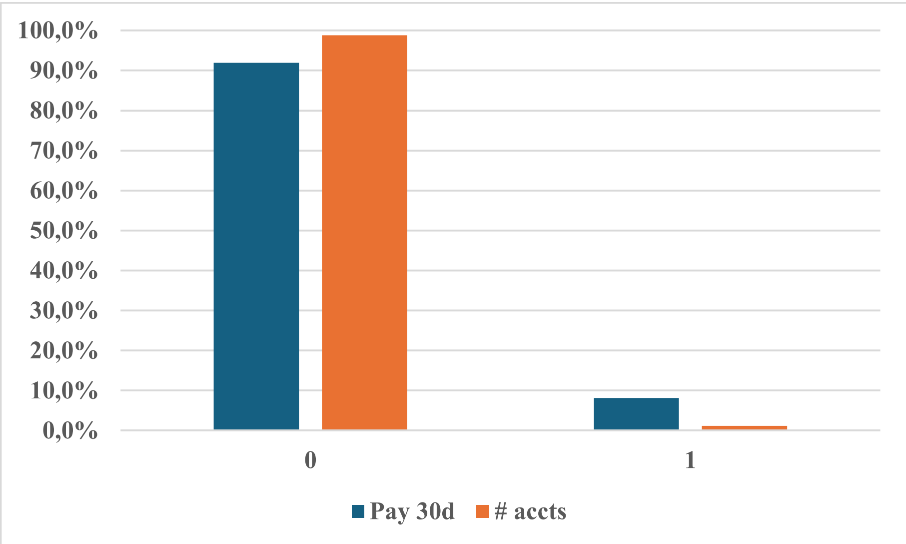
RPC last 90d flag	Pay 30d	# accts
0	75,2%	97,3%
1	24,8%	2,7%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 222 and Figure 225. Bivariant analysis RPC last 90d flag – payment 30 days for SMEs

RPC last 90d flag	Pay 30d	# accts
0	91,9%	98,8%
1	8,1%	1,2%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.28.- Variable "IN36_PTP _L90D" – Promise to pay in the last 90 days flag

Description

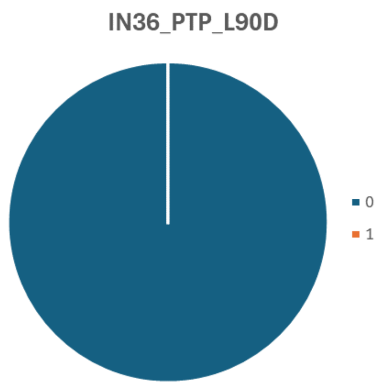
This variable reflects if the account has had a Promise to pay (PTP) in the last 90 days for every month of the analysis period. For the construction of this flag, it was used the last available data point for any given month. And this variable was included in the data set, as normally shows correlation with the probability to collect. The collections and recoveries actions (SMS, calls…) have a positive correlation with the probability to collect: with a higher number of collection actions, the probability to collect increases. Despite this factor, this variable was not relevant in our modelling.

Individuals' details

Analysing the last available data point for each bionomy month-account, we observe that only 20 of the total accounts had a promise to pay in the last 90 days.

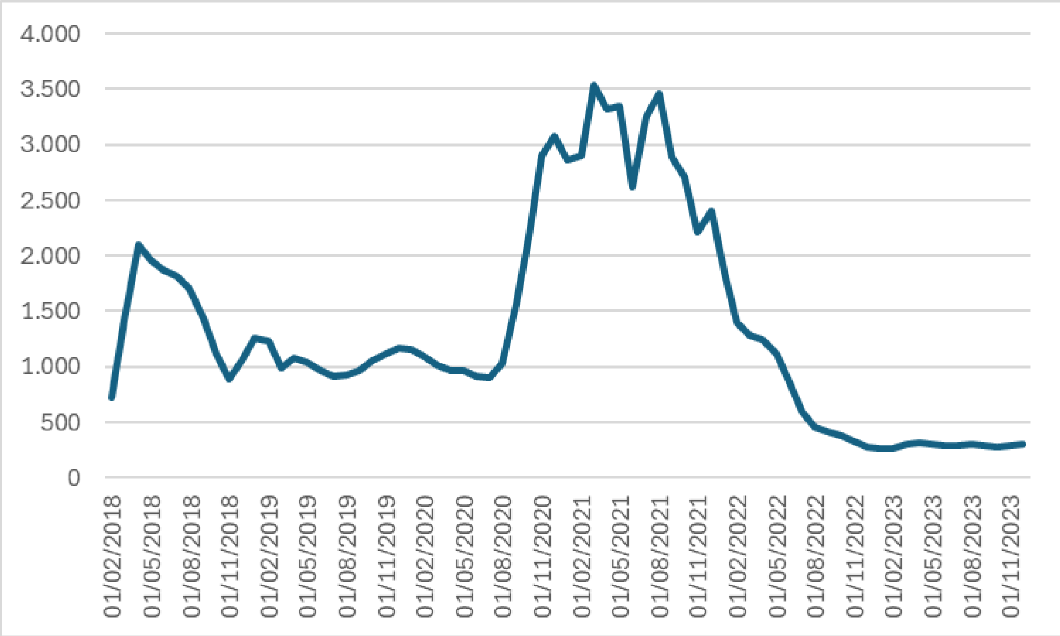
Table 223 and Figure 226. Distribution of Individuals accounts by PTP L90D flag

IN36_PTP_L90D	Record_Count
0	676.008
1	20



Source: Own elaboration

Figure 227. Monthly evolution of Individuals accounts by PTP L90D flag



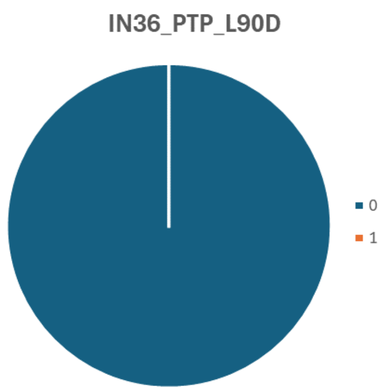
Source: Own elaboration

SMEs details

Analysing the last available data point for each bionomy month-account, we observe that none of the accounts had a promise to pay in the last 90 days.

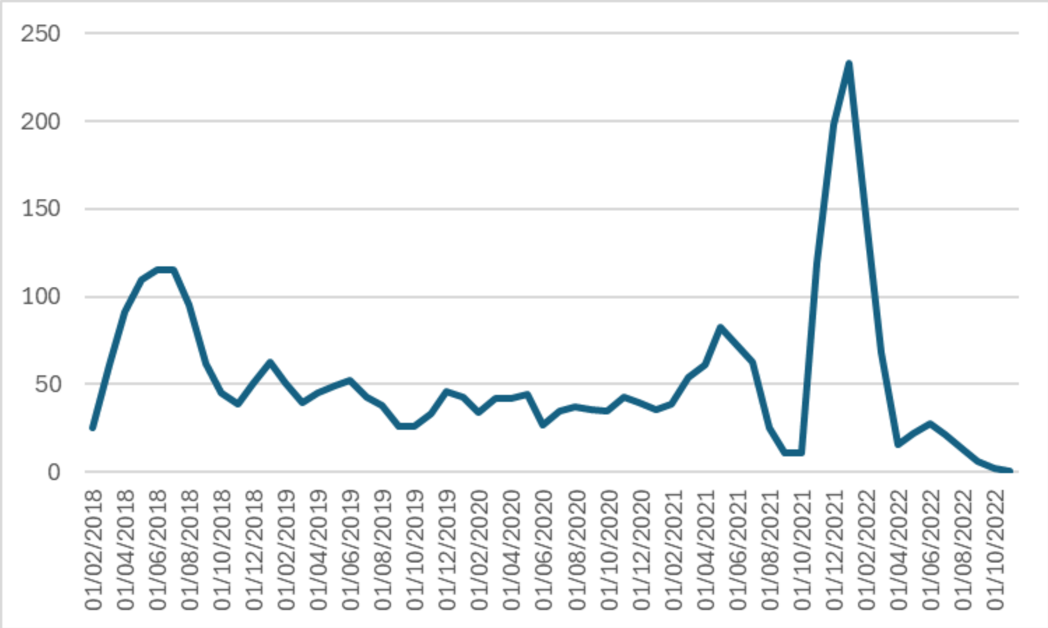
Table 224 and Figure 228. Distribution of SMEs accounts by PTP L90D flag

IN36_PTP_L90D	Record_Count
0	195.962
1	0



Source: Own elaboration

Figure 229. Monthly evolution of SMEs accounts by PTP L90D flag



* No new PTP after sep-22

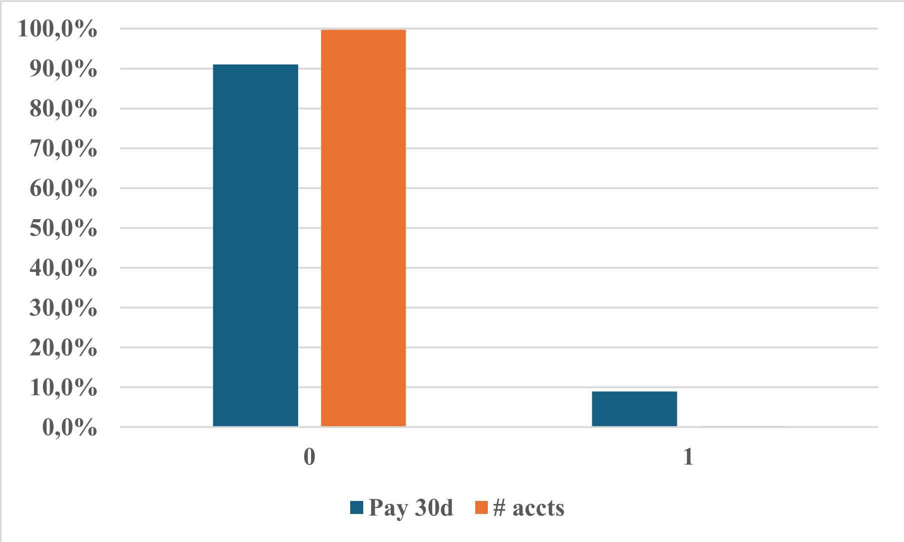
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

In this case, we observe that there is correlation between the PTP in the last 90 days flag and the payment in the following 30 days (the distribution of accounts and payments is not similar for the positive and negative results of the flag). Both in Individuals and SMEs, the probability for a payment is higher when the flag result is positive. The information value (IV) of this variable is 0.31 for Individuals and 0.02 for SMEs.

Table 225 and Figure 230. Bivariant analysis PTP last 90d flag – payment 30 days for Individuals

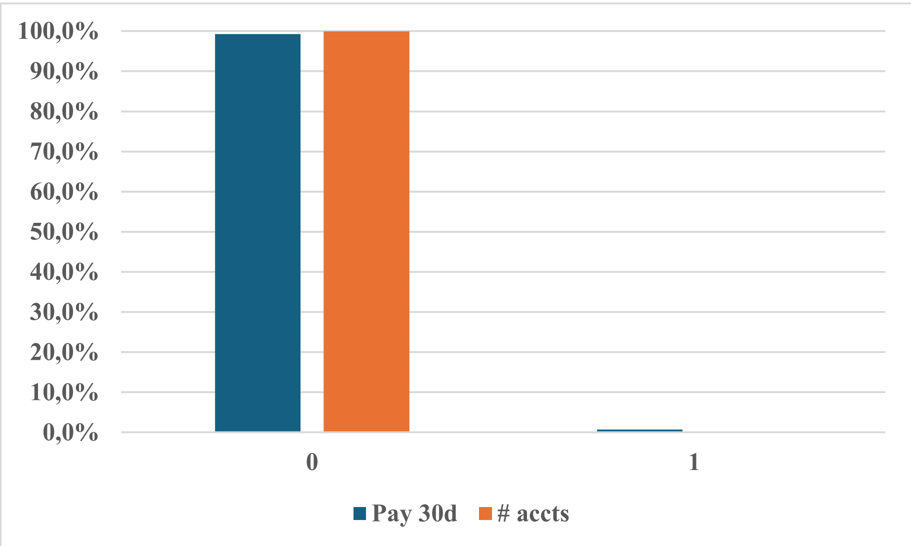
PTP last 90d flag	Pay 30d	# accts
0	91,0%	99,7%
1	9,0%	0,3%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 226 and Figure 231. Bivariant analysis PTP last 90d flag – payment 30 days for SMEs

PTP last 90d flag	Pay 30d	# accts
0	99,3%	100,0%
1	0,7%	0,0%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.29.- Variable “IN37_SMS_EVER” – SMS sent ever flag

Description

This variable reflects if the account has had an SMS sent ever for every month of the analysis period. For the construction of this flag, it was used the last available data point for any given month.

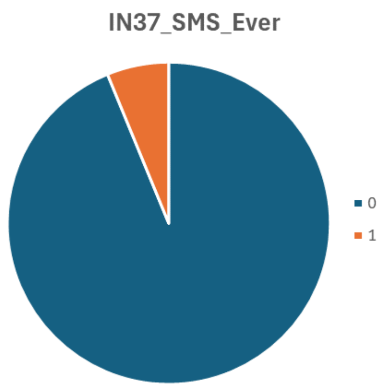
This variable was included in the data set, as normally shows correlation with the probability to collect. The collections and recoveries actions (SMS, calls…) have a positive correlation with the probability to collect: with a higher number of collection actions, the probability to collect increases. As explained in Section 4.4.3 of this document, this variable was eliminated for our modelling due to the excessive correlation with the variable to predict.

Individuals’ details

Analysing the last available data point for each bionomy month-account, we observe that 6.2% of the accounts have an SMS sent ever (in the analysis period).

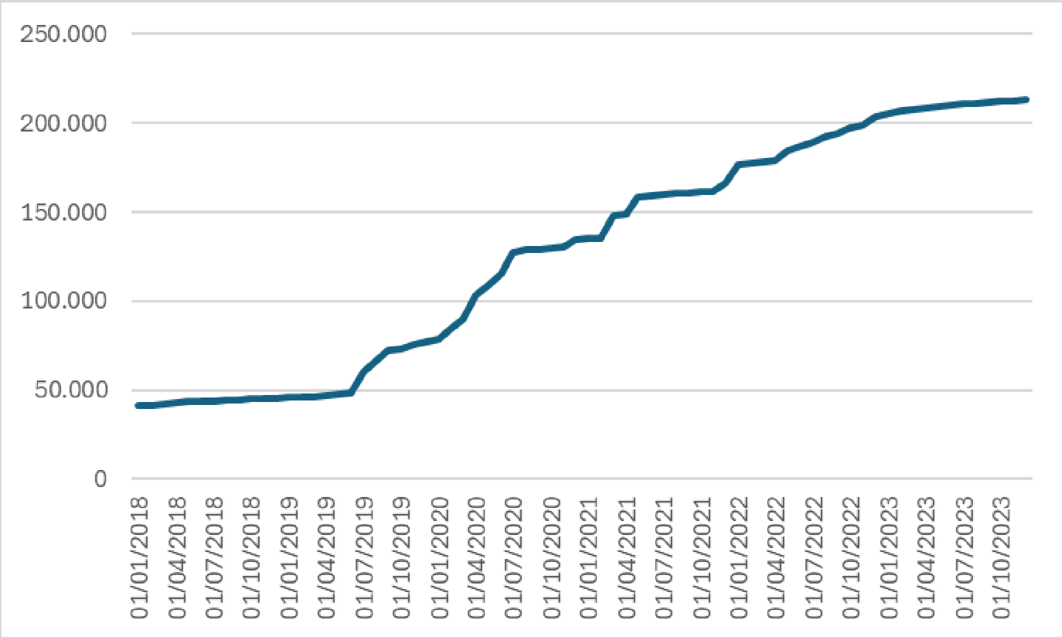
Table 227 and Figure 232. Distribution of Individuals accounts by SMS Ever flag

IN37_SMS_Ever	Record_Count
0	634.125
1	41.903



Source: Own elaboration

Figure 233. Monthly evolution of Individuals accounts by SMS Ever flag



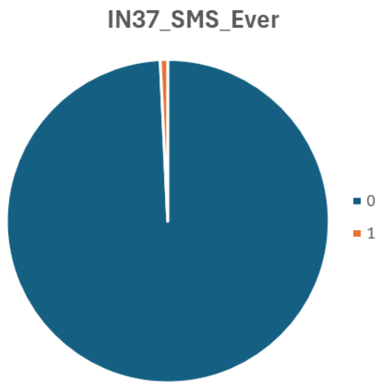
Source: Own elaboration

SMEs details

Analysing the last available data point for each bionomy month-account, we observe that just 0.7% of the accounts have an SMS sent ever (in the analysis period).

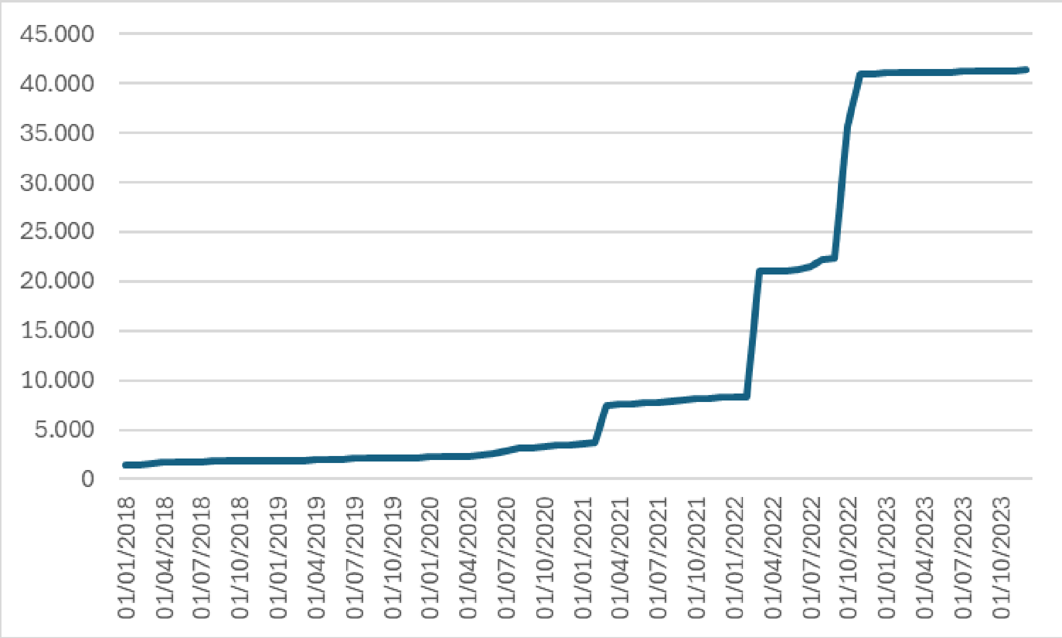
Table 228 and Figure 234. Distribution of SMEs accounts by SMS Ever flag

IN37_SMS_Ever	Record_Count
0	194.494
1	1.468



Source: Own elaboration

Figure 235. Monthly evolution of SMEs accounts by SMS Ever flag



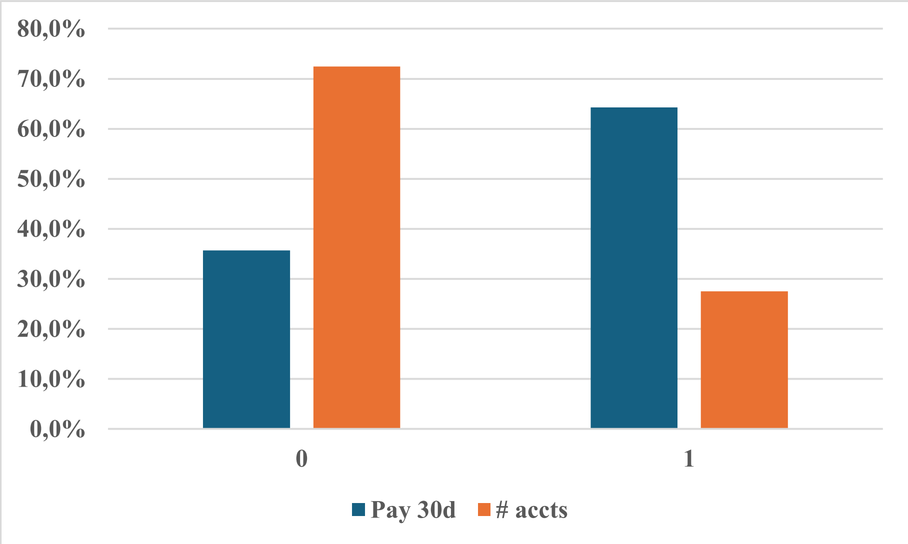
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

In this variable, we observe that there is an important correlation between the SMS sent ever flag and the payment in the following 30 days (the distribution of accounts and payments is not similar for the positive and negative results of the flag). Both in Individuals and SMEs, the probability for a payment is much higher when the flag result is positive. The information value (IV) of this variable is 0.57 for Individuals and 0.20 for SMEs.

Table 229 and Figure 236. Bivariant analysis SMS sent ever flag – payment 30 days for Individuals

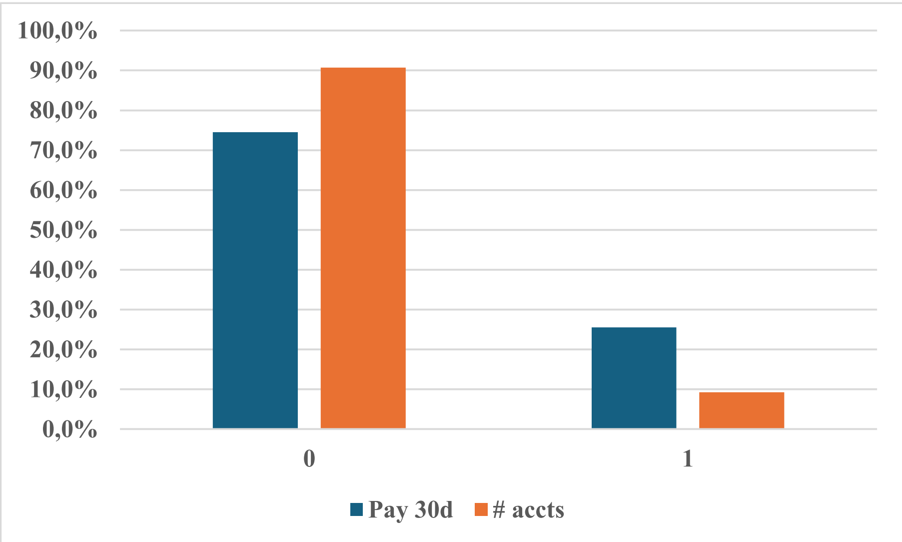
SMS sent ever flag	Pay 30d	# accts
0	35,7%	72,5%
1	64,3%	27,5%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 230 and Figure 237. Bivariant analysis SMS sent ever flag – payment 30 days for SMEs

SMS sent ever flag	Pay 30d	# accts
0	74,5%	90,8%
1	25,5%	9,2%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.30.- Variable "IN38_N_LLAMADAS_EVER" – Calls ever flag

Description

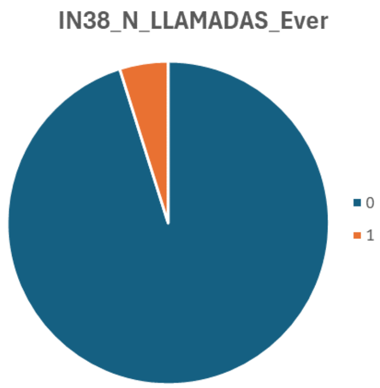
This variable reflects if the account has had a Call ever for every month of the analysis period. For the construction of this flag, it was used the last available data point for any given month. And this variable was included in the data set, as normally shows correlation with the probability to collect. The collections and recoveries actions (SMS, calls…) have a positive correlation with the probability to collect: with a higher number of collection actions, the probability to collect increases. Finally, this variable was relevant in our modelling for all the models, but the SMEs Covid-19 one.

Individuals' details

Analysing the last available data point for each bionomy month-account, we observe that just 4.9% of the accounts had a call ever (in the analysis period).

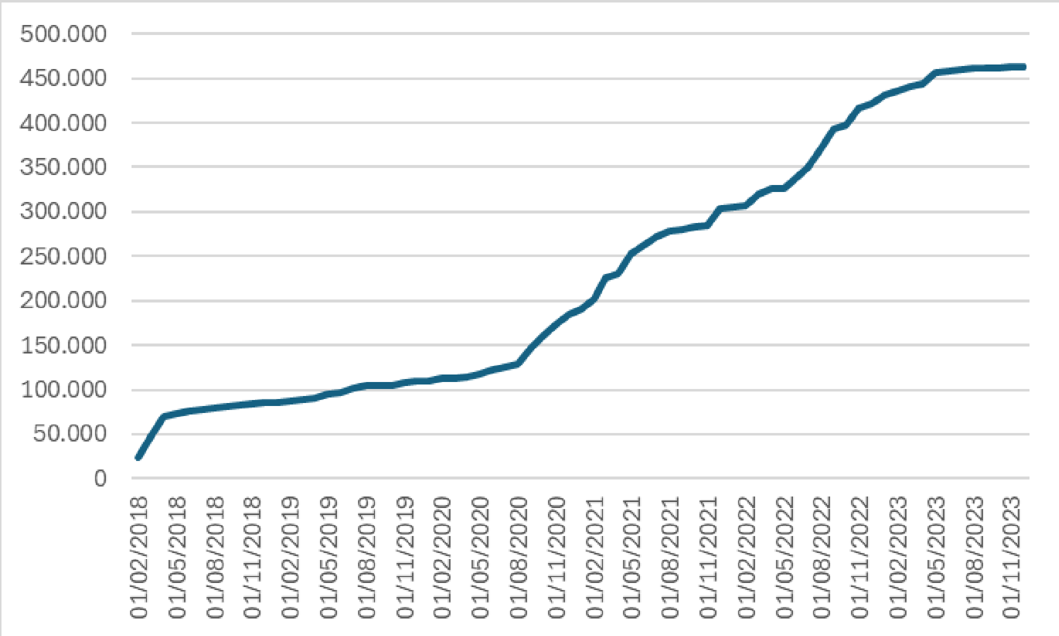
Table 231 and Figure 238. Distribution of Individuals accounts by Call Ever flag

IN38_N_LLAMADAS_Ever	Record_Count
0	643.095
1	32.933



Source: Own elaboration

Figure 239. Monthly evolution of Individuals accounts by Call Ever flag



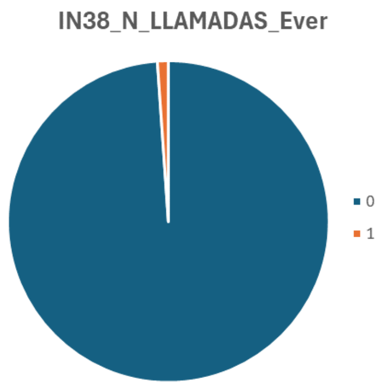
Source: Own elaboration

SMEs details

Analysing the last available data point for each bionomy month-account, we observe that just 1.1% of the accounts had a call ever (in the analysis period).

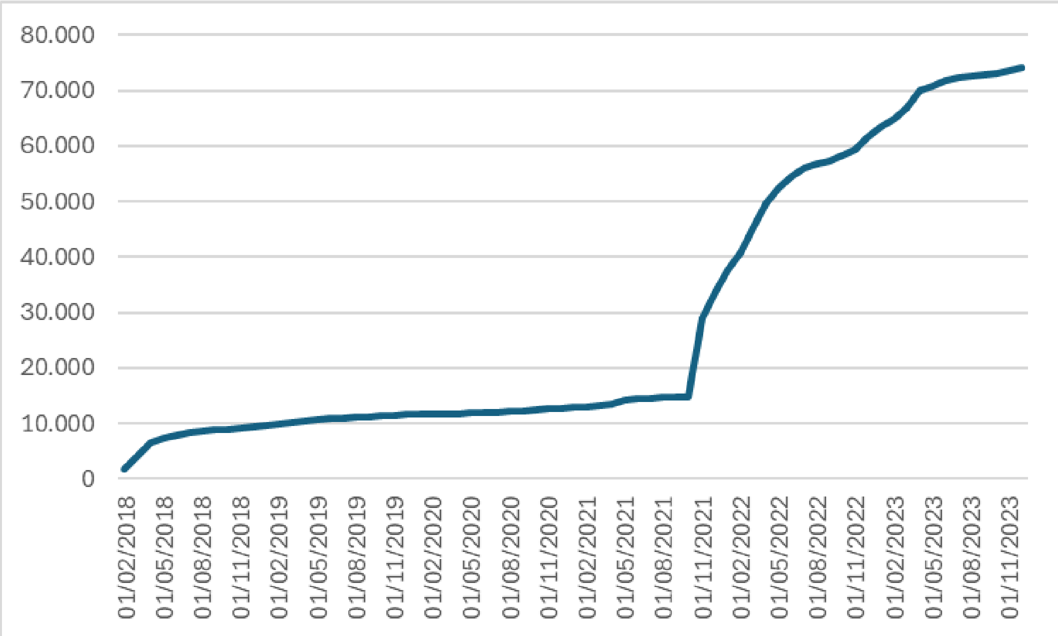
Table 232 and Figure 240. Distribution of SMEs accounts by Call Ever flag

IN38_N_LLAMADAS_Ever	Record_Count
0	193.814
1	2.148



Source: Own elaboration

Figure 241. Monthly evolution of SMEs accounts by Call Ever flag



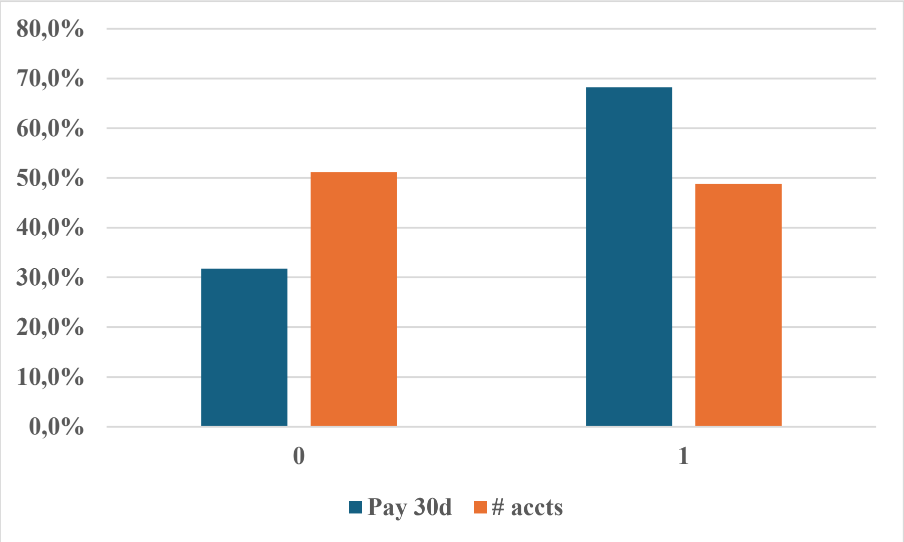
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

For this variable, we observe that there is correlation between the calls ever flag and the payment in the following 30 days (the distribution of accounts and payments is not similar for the positive and negative results of the flag). Both in Individuals and SMEs, the probability for a payment is higher when the flag result is positive. The information value (IV) of this variable is 0.16 for Individuals and 0.20 for SMEs.

Table 233 and Figure 242. Bivariant analysis calls ever flag – payment 30 days for Individuals

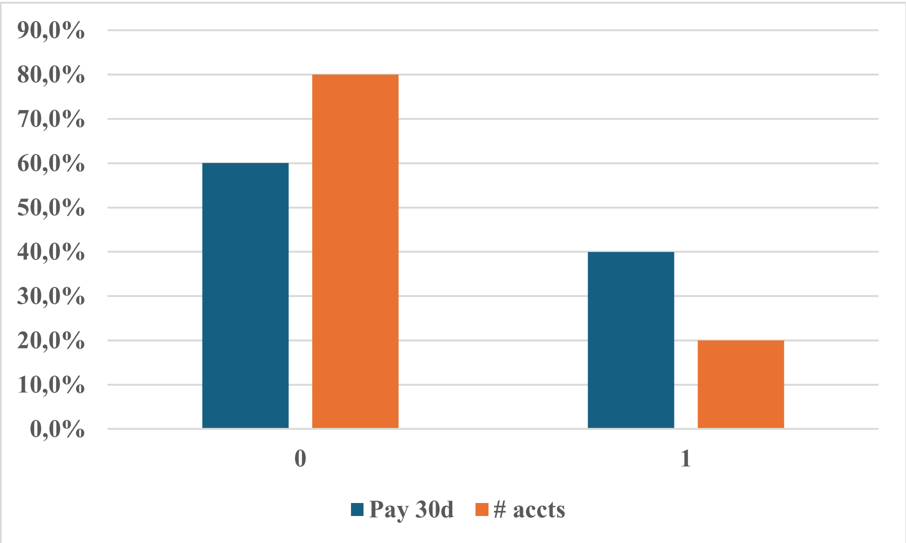
Calls ever flag	Pay 30d	# accts
0	31,7%	51,2%
1	68,3%	48,8%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 234 and Figure 243. Bivariant analysis calls ever flag – payment 30 days for SMEs

Calls ever flag	Pay 30d	# accts
0	60,1%	80,0%
1	39,9%	20,0%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

3.31.- Variable “IN39_RPC_EVER” – Right-party contact ever flag

Description

This variable reflects if the account has had a Right-party Contact (RPC) ever for every month of the analysis period. For the construction of this flag, it was used the last available data point for any given month.

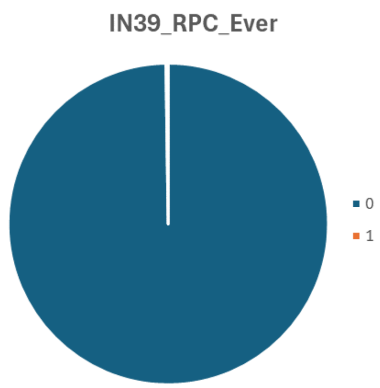
This variable was included in the data set, as normally shows correlation with the probability to collect. The collections and recoveries actions (SMS, calls…) have a positive correlation with the probability to collect: with a higher number of collection actions, the probability to collect increases. For the models built, this variable was relevant for the Individuals Covid-19, Individuals Post Covid-19 and SMEs Post Covid-19 ones.

Individuals’ details

Analysing the last available data point for each bionomy month-account, we observe that only 0.2% of the accounts had a right-party contact ever (in the analysis period).

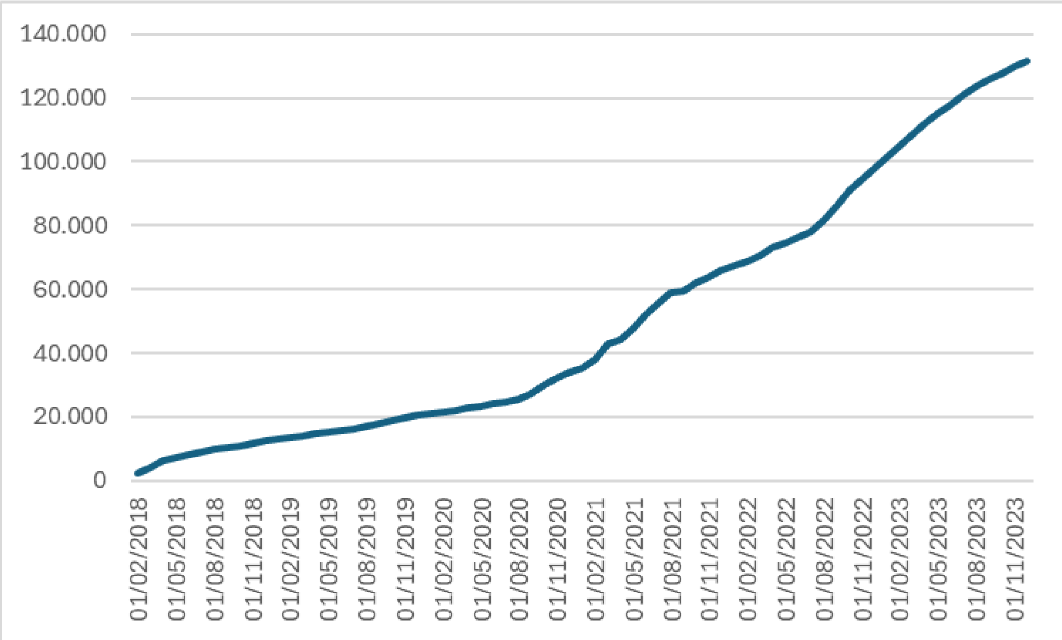
Table 235 and Figure 244. Distribution of Individuals accounts by RPC Ever flag

IN39_RPC_Ever	Record_Count
0	674.503
1	1.525



Source: Own elaboration

Figure 245. Monthly evolution of Individuals accounts by RPC Ever flag



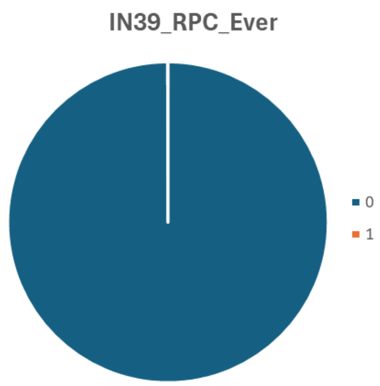
Source: Own elaboration

SMEs details

Analysing the last available data point for each bionomy month-account, we observe that only 0.04% of the accounts had a right-party contact ever (in the analysis period).

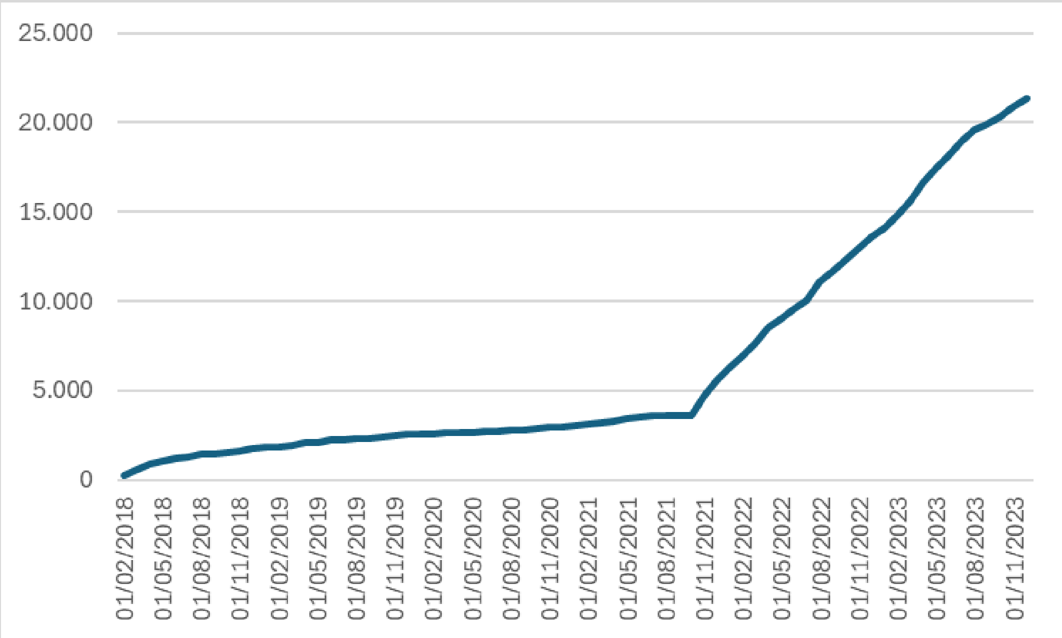
Table 236 and Figure 246. Distribution of SMEs accounts by RPC Ever flag

IN39_RPC_Ever	Record_Count
0	195.887
1	75



Source: Own elaboration

Figure 247. Monthly evolution of SMEs accounts by RPC Ever flag



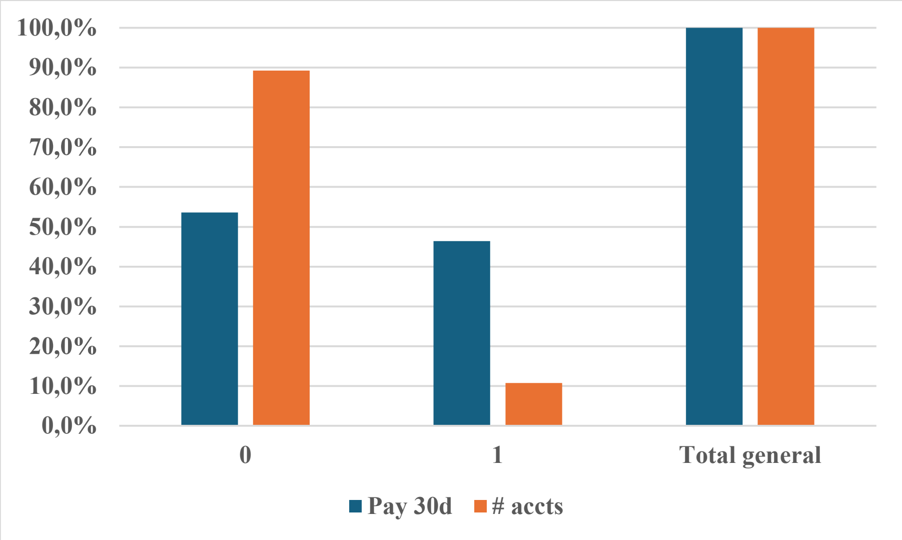
Source: Own elaboration

Discrimination bivariant analysis for Individuals and SMEs

For this variable, we observe that there is correlation between the RPC ever flag and the payment in the following 30 days (the distribution of accounts and payments is not similar for the positive and negative results of the flag). Both in Individuals and SMEs, the probability for a payment is higher when the flag result is positive. The information value (IV) of this variable is 0.70 for Individuals and 0.26 for SMEs.

Table 237 and Figure 248. Bivariant analysis RPC ever flag – payment 30 days for Individuals

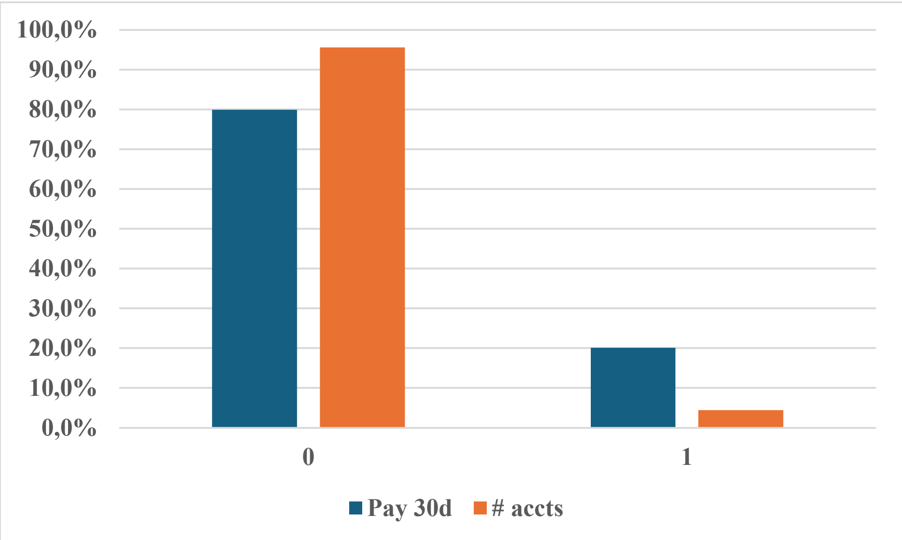
RPC ever flag	Pay 30d	# accts
0	53,6%	89,2%
1	46,4%	10,8%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

Table 238 and Figure 249. Bivariant analysis RPC ever flag – payment 30 days for SMEs

RPC ever flag	Pay 30d	# accts
0	79,9%	95,5%
1	20,1%	4,5%
Total general	**100,0%**	**100,0%**



Source: Own elaboration

APPENDIX 4.- MACROECONOMIC FACTORS DETAIL

4.1.- IPC General Var Anu

Table 239. IPC General Var Anu

Variable1	Valor1	Variable2	Valor2	Variable3	Valor3	PERIODO	VALOR	Fecha
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2023M07	2,3	01/07/2023
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2023M06	1,9	01/06/2023
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2023M05	3,2	01/05/2023
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2023M04	4,1	01/04/2023
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2023M03	3,3	01/03/2023
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2023M02	6,0	01/02/2023
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2023M01	5,9	01/01/2023
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M12	5,7	01/12/2022
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M11	6,8	01/11/2022
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M10	7,3	01/10/2022
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M09	8,9	01/09/2022
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M08	10,5	01/08/2022
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M07	10,8	01/07/2022
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M06	10,2	01/06/2022
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M05	8,7	01/05/2022
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M04	8,3	01/04/2022
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M03	9,8	01/03/2022
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M02	7,6	01/02/2022
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M01	6,1	01/01/2022
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M12	6,5	01/12/2021
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M11	5,5	01/11/2021
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M10	5,4	01/10/2021
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M09	4,0	01/09/2021
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M08	3,3	01/08/2021
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M07	2,9	01/07/2021
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M06	2,7	01/06/2021
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M05	2,7	01/05/2021
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M04	2,2	01/04/2021
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M03	1,3	01/03/2021
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M02	0,0	01/02/2021
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M01	0,5	01/01/2021
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M12	-0,5	01/12/2020
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M11	-0,8	01/11/2020
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M10	-0,8	01/10/2020
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M09	-0,4	01/09/2020
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M08	-0,5	01/08/2020
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M07	-0,6	01/07/2020
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M06	-0,3	01/06/2020
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M05	-0,9	01/05/2020
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M04	-0,7	01/04/2020
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M03	0,0	01/03/2020
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M02	0,7	01/02/2020
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M01	1,1	01/01/2020
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M12	0,8	01/12/2019
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M11	0,4	01/11/2019
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M10	0,1	01/10/2019
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M09	0,1	01/09/2019
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M08	0,3	01/08/2019
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M07	0,5	01/07/2019
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M06	0,4	01/06/2019
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M05	0,8	01/05/2019
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M04	1,5	01/04/2019
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M03	1,3	01/03/2019
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M02	1,1	01/02/2019
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M01	1,0	01/01/2019
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M12	1,2	01/12/2018
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M11	1,7	01/11/2018
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M10	2,3	01/10/2018
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M09	2,3	01/09/2018
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M08	2,2	01/08/2018
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M07	2,2	01/07/2018
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M06	2,3	01/06/2018
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M05	2,1	01/05/2018
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M04	1,1	01/04/2018
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M03	1,2	01/03/2018
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M02	1,1	01/02/2018
Tipo de dato	Variación anual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M01	0,6	01/01/2018

Source: INE

4.2.- IPC General Var Mens

Table 240. IPC General Var Mens

Variable1	Valor1	Variable2	Valor2	Variable3	Valor3	PERIODO	VALOR	Fecha
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2023M07	0,1	01/07/2023
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2023M06	0,6	01/06/2023
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2023M05	0	01/05/2023
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2023M04	0,6	01/04/2023
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2023M03	0,4	01/03/2023
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2023M02	0,9	01/02/2023
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2023M01	-0,2	01/01/2023
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M12	0,2	01/12/2022
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M11	-0,1	01/11/2022
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M10	0,3	01/10/2022
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M09	-0,7	01/09/2022
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M08	0,3	01/08/2022
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M07	-0,3	01/07/2022
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M06	1,9	01/06/2022
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M05	0,8	01/05/2022
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M04	-0,2	01/04/2022
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M03	3	01/03/2022
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M02	0,8	01/02/2022
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2022M01	-0,4	01/01/2022
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M12	1,2	01/12/2021
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M11	0,3	01/11/2021
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M10	1,8	01/10/2021
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M09	0,8	01/09/2021
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M08	0,5	01/08/2021
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M07	-0,8	01/07/2021
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M06	0,5	01/06/2021
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M05	0,5	01/05/2021
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M04	1,2	01/04/2021
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M03	1	01/03/2021
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M02	-0,6	01/02/2021
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2021M01	0	01/01/2021
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M12	0,2	01/12/2020
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M11	0,2	01/11/2020
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M10	0,5	01/10/2020
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M09	0,2	01/09/2020
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M08	0	01/08/2020
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M07	-0,9	01/07/2020
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M06	0,5	01/06/2020
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M05	0	01/05/2020
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M04	0,3	01/04/2020
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M03	-0,4	01/03/2020
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M02	-0,1	01/02/2020
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2020M01	-1	01/01/2020
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M12	-0,1	01/12/2019
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M11	0,2	01/11/2019
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M10	1	01/10/2019
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M09	0	01/09/2019
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M08	-0,1	01/08/2019
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M07	-0,6	01/07/2019
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M06	-0,1	01/06/2019
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M05	0,2	01/05/2019
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M04	1	01/04/2019
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M03	0,4	01/03/2019
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M02	0,2	01/02/2019
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2019M01	-1,3	01/01/2019
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M12	-0,4	01/12/2018
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M11	-0,1	01/11/2018
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M10	0,9	01/10/2018
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M09	0,2	01/09/2018
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M08	0,1	01/08/2018
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M07	-0,7	01/07/2018
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M06	0,3	01/06/2018
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M05	0,9	01/05/2018
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M04	0,8	01/04/2018
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M03	0,1	01/03/2018
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M02	0,1	01/02/2018
Tipo de dato	Variación mensual	Total Nacional	Total Nacional	Grupos ECOICOP	Índice general	2018M01	-1,1	01/01/2018

Source: INE

4.3.- IPC Subyac Var Anu

Table 241. IPC Subyac Var Anu

Variable1	Valor1	Variable2	Valor2	Variable3	Valor3	PERIODO	VALOR	Fecha
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2023M07	6,2	01/07/2023
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2023M06	5,9	01/06/2023
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2023M05	6,1	01/05/2023
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2023M04	6,6	01/04/2023
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2023M03	7,5	01/03/2023
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2023M02	7,6	01/02/2023
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2023M01	7,5	01/01/2023
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M12	7,0	01/12/2022
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M11	6,3	01/11/2022
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M10	6,2	01/10/2022
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M09	6,2	01/09/2022
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M08	6,4	01/08/2022
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M07	6,1	01/07/2022
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M06	5,5	01/06/2022
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M05	4,9	01/05/2022
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M04	4,4	01/04/2022
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M03	3,4	01/03/2022
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M02	3,0	01/02/2022
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M01	2,4	01/01/2022
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M12	2,1	01/12/2021
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M11	1,7	01/11/2021
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M10	1,4	01/10/2021
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M09	1,0	01/09/2021
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M08	0,7	01/08/2021
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M07	0,6	01/07/2021
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M06	0,2	01/06/2021
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M05	0,2	01/05/2021
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M04	0,0	01/04/2021
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M03	0,3	01/03/2021
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M02	0,3	01/02/2021
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M01	0,6	01/01/2021
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M12	0,1	01/12/2020
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M11	0,2	01/11/2020
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M10	0,3	01/10/2020
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M09	0,4	01/09/2020
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M08	0,4	01/08/2020
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M07	0,6	01/07/2020
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M06	1,0	01/06/2020
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M05	1,1	01/05/2020
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M04	1,1	01/04/2020
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M03	1,1	01/03/2020
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M02	1,2	01/02/2020
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M01	1,0	01/01/2020
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M12	1,0	01/12/2019
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M11	1,0	01/11/2019
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M10	1,0	01/10/2019
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M09	1,0	01/09/2019
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M08	0,9	01/08/2019
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M07	0,9	01/07/2019
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M06	0,9	01/06/2019
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M05	0,7	01/05/2019
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M04	0,9	01/04/2019
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M03	0,7	01/03/2019
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M02	0,7	01/02/2019
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M01	0,8	01/01/2019
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M12	0,9	01/12/2018
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M11	0,9	01/11/2018
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M10	1,0	01/10/2018
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M09	0,8	01/09/2018
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M08	0,8	01/08/2018
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M07	0,9	01/07/2018
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M06	1,0	01/06/2018
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M05	1,1	01/05/2018
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M04	0,8	01/04/2018
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M03	1,2	01/03/2018
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M02	1,1	01/02/2018
Tipo de dato	Variación anual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M01	0,8	01/01/2018

Source: INE

4.4.- IPC Subyac Var Mens

Table 242. IPC Subyac Var Mens

Variable1	Valor1	Variable2	Valor2	Variable3	Valor3	PERIODO	VALOR	Fecha
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2023M07	0,0	01/07/2023
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2023M06	0,5	01/06/2023
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2023M05	0,2	01/05/2023
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2023M04	1,0	01/04/2023
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2023M03	0,7	01/03/2023
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2023M02	0,7	01/02/2023
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2023M01	-0,2	01/01/2023
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M12	0,9	01/12/2022
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M11	0,7	01/11/2022
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M10	1,0	01/10/2022
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M09	0,2	01/09/2022
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M08	0,4	01/08/2022
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M07	-0,3	01/07/2022
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M06	0,7	01/06/2022
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M05	0,7	01/05/2022
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M04	1,8	01/04/2022
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M03	0,8	01/03/2022
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M02	0,6	01/02/2022
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2022M01	-0,7	01/01/2022
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M12	0,3	01/12/2021
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M11	0,6	01/11/2021
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M10	1,0	01/10/2021
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M09	0,3	01/09/2021
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M08	0,1	01/08/2021
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M07	-0,9	01/07/2021
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M06	0,1	01/06/2021
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M05	0,3	01/05/2021
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M04	0,9	01/04/2021
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M03	0,4	01/03/2021
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M02	0,0	01/02/2021
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2021M01	-1,0	01/01/2021
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M12	0,0	01/12/2020
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M11	0,3	01/11/2020
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M10	0,6	01/10/2020
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M09	0,1	01/09/2020
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M08	0,0	01/08/2020
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M07	-1,2	01/07/2020
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M06	0,1	01/06/2020
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M05	0,1	01/05/2020
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M04	1,1	01/04/2020
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M03	0,5	01/03/2020
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M02	0,3	01/02/2020
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2020M01	-1,5	01/01/2020
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M12	0,0	01/12/2019
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M11	0,3	01/11/2019
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M10	0,8	01/10/2019
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M09	0,1	01/09/2019
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M08	0,2	01/08/2019
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M07	-0,8	01/07/2019
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M06	0,2	01/06/2019
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M05	0,1	01/05/2019
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M04	1,1	01/04/2019
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M03	0,5	01/03/2019
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M02	0,1	01/02/2019
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2019M01	-1,5	01/01/2019
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M12	0,0	01/12/2018
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M11	0,2	01/11/2018
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M10	0,8	01/10/2018
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M09	0,0	01/09/2018
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M08	0,1	01/08/2018
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M07	-0,8	01/07/2018
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M06	0,1	01/06/2018
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M05	0,4	01/05/2018
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M04	0,8	01/04/2018
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M03	0,5	01/03/2018
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M02	0,3	01/02/2018
Tipo de dato	Variación mensual	Grupos especiales 2001	General sin alimentos no elaborados ni productos energéticos	Total Nacional	Total Nacional	2018M01	-1,4	01/01/2018

Source: INE

4.5.- ICC

Table 243. ICC

Mes	ICC	Ind_Sit_Actual	Ind_Expectativas
01/01/2018	102,3	93,5	111,1
01/02/2018	99,7	90,6	108,7
01/03/2018	98,6	90,1	107,2
01/04/2018	99,9	92,7	107,2
01/05/2018	97,7	91,3	104,1
01/06/2018	107,0	102,0	112,0
01/07/2018	106,1	101,7	110,4
01/08/2018	102,4	99,3	105,6
01/09/2018	90,6	82,9	98,2
01/10/2018	93,0	85,4	100,5
01/11/2018	91,4	83,8	99,0
01/12/2018	90,9	83,4	98,4
01/01/2019	93,8	87,7	99,8
01/02/2019	96,1	86,8	105,5
01/03/2019	93,9	85,2	102,6
01/04/2019	97,0	88,5	105,5
01/05/2019	96,9	89,3	104,4
01/06/2019	102,3	96,0	108,6
01/07/2019	97,0	93,2	100,9
01/08/2019	86,0	83,1	88,9
01/09/2019	80,7	76,7	84,7
01/10/2019	73,3	70,0	76,6
01/11/2019	77,4	69,5	85,2
01/12/2019	77,7	70,1	85,4
01/01/2020	87,2	76,6	97,7
01/02/2020	85,7	75,9	95,4
01/03/2020	63,3	57,2	69,4
01/04/2020	49,9	31,5	68,3
01/05/2020	52,9	27,6	78,2
01/06/2020	60,7	30,9	90,5
01/07/2020	53,1	30,1	76,1
01/08/2020	49,9	29,7	70,1
01/09/2020	49,5	29,2	69,7
01/10/2020	48,5	28,2	68,8
01/11/2020	55,7	29,9	81,5
01/12/2020	63,1	34,2	92,0
01/01/2021	55,7	32,9	78,6
01/02/2021	65,9	37,2	94,6
01/03/2021	73,0	43,4	102,7
01/04/2021	77,8	48,1	107,4
01/05/2021	89,0	62,4	115,5
01/06/2021	97,5	75,9	119,1
01/07/2021	91,9	73,6	110,3
01/08/2021	91,6	77,9	105,2
01/09/2021	98,3	83,1	113,4
01/10/2021	97,3	85,0	109,6
01/11/2021	84,6	74,3	94,9
01/12/2021	81,3	69,8	92,8
01/01/2022	89,3	73,6	105,1
01/02/2022	89,8	80,9	98,6
01/03/2022	53,8	49,5	58,1
01/04/2022	74,6	61,1	88,0
01/05/2022	76,0	66,1	85,9
01/06/2022	65,8	59,5	72,2
01/07/2022	55,5	53,1	57,9
01/08/2022	55,5	53,1	57,9
01/09/2022	55,7	47,2	64,3
01/10/2022	54,7	46,3	63,1
01/11/2022	60,5	49,8	71,2
01/12/2022	68,0	56,1	80,0
01/01/2023	73,0	60,2	85,8
01/02/2023	71,6	59,2	84,0
01/03/2023	67,4	54,8	80,1
01/04/2023	73,0	62,8	83,1
01/05/2023	81,5	71,5	91,5
01/06/2023	92,4	83,2	101,6

Source: INE

4.6.- Tasa Actv Var Anu

Table 244. Tasa Actv Var Anu

Variable1	Valor1	Variable2	Valor2	Variable3	Valor3	Variable4	Valor4	Variable5	Valor5	PERIODO	VALOR
Tipo de dato	Tasa de actividad	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2023T2	0,3
Tipo de dato	Tasa de actividad	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2023T1	0,1
Tipo de dato	Tasa de actividad	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2022T4	-0,1
Tipo de dato	Tasa de actividad	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2022T3	-0,3
Tipo de dato	Tasa de actividad	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2022T2	0,1
Tipo de dato	Tasa de actividad	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2022T1	0,8
Tipo de dato	Tasa de actividad	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2021T4	0,5
Tipo de dato	Tasa de actividad	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2021T3	1,3
Tipo de dato	Tasa de actividad	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2021T2	3,0
Tipo de dato	Tasa de actividad	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2021T1	-0,5
Tipo de dato	Tasa de actividad	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2020T4	-0,6
Tipo de dato	Tasa de actividad	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2020T3	-0,9
Tipo de dato	Tasa de actividad	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2020T2	-3,2
Tipo de dato	Tasa de actividad	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2020T1	-0,2
Tipo de dato	Tasa de actividad	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2019T4	0,1
Tipo de dato	Tasa de actividad	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2019T3	0,0
Tipo de dato	Tasa de actividad	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2019T2	-0,1
Tipo de dato	Tasa de actividad	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2019T1	-0,1
Tipo de dato	Tasa de actividad	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2018T4	-0,2
Tipo de dato	Tasa de actividad	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2018T3	-0,2
Tipo de dato	Tasa de actividad	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2018T2	0,0
Tipo de dato	Tasa de actividad	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2018T1	-0,3

Source: INE

4.7.- Tasa Empleo Var Anu

Table 245. Tasa Empleo Var Anu

Variable1	Valor1	Variable2	Valor2	Variable3	Valor3	Variable4	Valor4	Variable5	Valor5	PERIODO	VALOR
Tipo de dato	Tasa de paro de la población	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2023T2	-0,9
Tipo de dato	Tasa de paro de la población	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2023T1	-0,4
Tipo de dato	Tasa de paro de la población	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2022T4	-0,5
Tipo de dato	Tasa de paro de la población	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2022T3	-1,9
Tipo de dato	Tasa de paro de la población	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2022T2	-2,8
Tipo de dato	Tasa de paro de la población	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2022T1	-2,3
Tipo de dato	Tasa de paro de la población	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2021T4	-2,8
Tipo de dato	Tasa de paro de la población	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2021T3	-1,7
Tipo de dato	Tasa de paro de la población	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2021T2	-0,1
Tipo de dato	Tasa de paro de la población	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2021T1	1,6
Tipo de dato	Tasa de paro de la población	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2020T4	2,4
Tipo de dato	Tasa de paro de la población	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2020T3	2,3
Tipo de dato	Tasa de paro de la población	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2020T2	1,3
Tipo de dato	Tasa de paro de la población	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2020T1	-0,3
Tipo de dato	Tasa de paro de la población	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2019T4	-0,7
Tipo de dato	Tasa de paro de la población	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2019T3	-0,6
Tipo de dato	Tasa de paro de la población	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2019T2	-1,3
Tipo de dato	Tasa de paro de la población	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2019T1	-2,1
Tipo de dato	Tasa de paro de la población	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2018T4	-2,1
Tipo de dato	Tasa de paro de la población	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2018T3	-1,8
Tipo de dato	Tasa de paro de la población	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2018T2	-1,9
Tipo de dato	Tasa de paro de la población	Sexo	Ambos sexos	Total Nacional	Total Nacional	Semiintervalos de edad	16 y más años	Tipo de dato: Tasas	Diferencia anual	2018T1	-2,0

Source: INE

4.8.- PIB equival ANNUAL

Table 246. PIB equival ANNUAL

Variable1	Valor1	Variable2	Valor2	Variable3	Valor3	Variable4	Valor4	Variable5	Valor5	Variable6	Valor6	PERIODO	VALOR
Tipo de dato	Variación anual	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2023T2	2,8774
Tipo de dato	Variación anual	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2023T1	2,2872
Tipo de dato	Variación anual	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2022T4	2,077
Tipo de dato	Variación anual	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2022T3	2,8257
Tipo de dato	Variación anual	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2022T2	5,0086
Tipo de dato	Variación anual	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2022T1	5,3217
Tipo de dato	Variación anual	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2021T4	5,9994
Tipo de dato	Variación anual	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2021T3	6,3638
Tipo de dato	Variación anual	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2021T2	18,9059
Tipo de dato	Variación anual	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2021T1	-2,7479
Tipo de dato	Variación anual	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2020T4	-4,6928
Tipo de dato	Variación anual	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2020T3	-5,1347
Tipo de dato	Variación anual	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2020T2	-17,7799
Tipo de dato	Variación anual	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2020T1	0,2661
Tipo de dato	Variación anual	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2019T4	3,1863
Tipo de dato	Variación anual	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2019T3	2,8519
Tipo de dato	Variación anual	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2019T2	3,6084
Tipo de dato	Variación anual	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2019T1	3,617
Tipo de dato	Variación anual	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2018T4	2,2299
Tipo de dato	Variación anual	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2018T3	2,2067
Tipo de dato	Variación anual	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2018T2	2,1324
Tipo de dato	Variación anual	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2018T1	2,2507

Source: INE

4.9.- PIB equival TRIMES

Table 247. PIB equival TRIMES

Variable1	Valor1	Variable2	Valor2	Variable3	Valor3	Variable4	Valor4	Variable5	Valor5	Variable6	Valor6	PERIODO	VALOR
Tipo de dato	Variación trimestral	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2023T2	1,3059
Tipo de dato	Variación trimestral	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2023T1	0,2096
Tipo de dato	Variación trimestral	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2022T4	0,1766
Tipo de dato	Variación trimestral	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2022T3	1,1601
Tipo de dato	Variación trimestral	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2022T2	0,7247
Tipo de dato	Variación trimestral	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2022T1	0,0037
Tipo de dato	Variación trimestral	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2021T4	0,9114
Tipo de dato	Variación trimestral	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2021T3	3,3077
Tipo de dato	Variación trimestral	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2021T2	1,025
Tipo de dato	Variación trimestral	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2021T1	0,6472
Tipo de dato	Variación trimestral	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2020T4	1,2583
Tipo de dato	Variación trimestral	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2020T3	15,4895
Tipo de dato	Variación trimestral	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2020T2	-17,3725
Tipo de dato	Variación trimestral	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2020T1	-1,3656
Tipo de dato	Variación trimestral	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2019T4	0,7888
Tipo de dato	Variación trimestral	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2019T3	0,0952
Tipo de dato	Variación trimestral	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2019T2	0,7628
Tipo de dato	Variación trimestral	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2019T1	1,5071
Tipo de dato	Variación trimestral	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2018T4	0,4621
Tipo de dato	Variación trimestral	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2018T3	0,8315
Tipo de dato	Variación trimestral	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2018T2	0,7712
Tipo de dato	Variación trimestral	Agregados de empleo	Puestos de trabajo equivalentes a tiempo completo	Total Nacional	Total Nacional	RELACIÓN CON LA ACTIVIDAD	Ocupado/a	TOTALES VARIABLES CNAE	Total CNAE	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2018T1	0,1482

Source: INE

4.10.- PIB mdo ANNUAL

Table 248. PIB mdo ANNUAL

Variable1	Valor1	Variable2	Valor2	Variable3	Valor3	Variable4	Valor4	Variable5	Valor5	PERIODO	VALOR
Tipo de dato	Variación anual	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2023T2	1,7532
Tipo de dato	Variación anual	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2023T1	4,1562
Tipo de dato	Variación anual	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2022T4	3,0225
Tipo de dato	Variación anual	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2022T3	4,9218
Tipo de dato	Variación anual	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2022T2	7,7656
Tipo de dato	Variación anual	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2022T1	6,2564
Tipo de dato	Variación anual	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2021T4	6,6253
Tipo de dato	Variación anual	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2021T3	4,202
Tipo de dato	Variación anual	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2021T2	17,8753
Tipo de dato	Variación anual	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2021T1	-4,43
Tipo de dato	Variación anual	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2020T4	-9,4668
Tipo de dato	Variación anual	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2020T3	-9,2169
Tipo de dato	Variación anual	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2020T2	-21,9369
Tipo de dato	Variación anual	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2020T1	-4,6736
Tipo de dato	Variación anual	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2019T4	1,4748
Tipo de dato	Variación anual	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2019T3	1,9226
Tipo de dato	Variación anual	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2019T2	2,1515
Tipo de dato	Variación anual	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2019T1	2,3956
Tipo de dato	Variación anual	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2018T4	2,2105
Tipo de dato	Variación anual	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2018T3	2,0886
Tipo de dato	Variación anual	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2018T2	2,1996
Tipo de dato	Variación anual	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2018T1	2,645

Source: INE

4.11.- PIB mdo TRIMES

Table 249. PIB mdo TRIMES

Variable1	Valor1	Variable2	Valor2	Variable3	Valor3	Variable4	Valor4	Variable5	Valor5	PERIODO	VALOR
Tipo de dato	Variación trimestral	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2023T2	0,4205
Tipo de dato	Variación trimestral	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2023T1	0,5201
Tipo de dato	Variación trimestral	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2022T4	0,4142
Tipo de dato	Variación trimestral	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2022T3	0,3870
Tipo de dato	Variación trimestral	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2022T2	2,7920
Tipo de dato	Variación trimestral	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2022T1	-0,5740
Tipo de dato	Variación trimestral	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2021T4	2,2654
Tipo de dato	Variación trimestral	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2021T3	3,1079
Tipo de dato	Variación trimestral	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2021T2	1,3524
Tipo de dato	Variación trimestral	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2021T1	-0,2288
Tipo de dato	Variación trimestral	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2020T4	-0,0587
Tipo de dato	Variación trimestral	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2020T3	16,6377
Tipo de dato	Variación trimestral	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2020T2	-17,8263
Tipo de dato	Variación trimestral	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2020T1	-5,4871
Tipo de dato	Variación trimestral	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2019T4	0,2171
Tipo de dato	Variación trimestral	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2019T3	0,2951
Tipo de dato	Variación trimestral	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2019T2	0,3461
Tipo de dato	Variación trimestral	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2019T1	0,6089
Tipo de dato	Variación trimestral	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2018T4	0,6594
Tipo de dato	Variación trimestral	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2018T3	0,5204
Tipo de dato	Variación trimestral	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2018T2	0,5859
Tipo de dato	Variación trimestral	Total Nacional	Total Nacional	SALDOS CONTABLES	Producto interior bruto a precios de mercado	Precios_SEC	Índices de volumen encadenados	Corrección de efectos	Datos ajustados de estacionalidad y calendario	2018T1	0,4270

Source: INE

APPENDIX 5.- MODELS DETAIL

5.1.- Complete models detail

5.1.1.- Individuals Global model (complete)

Variables included in the model and their importance as predictor

The variables included in this model are the following 25 (ordered in descendent importance) in Table 250.

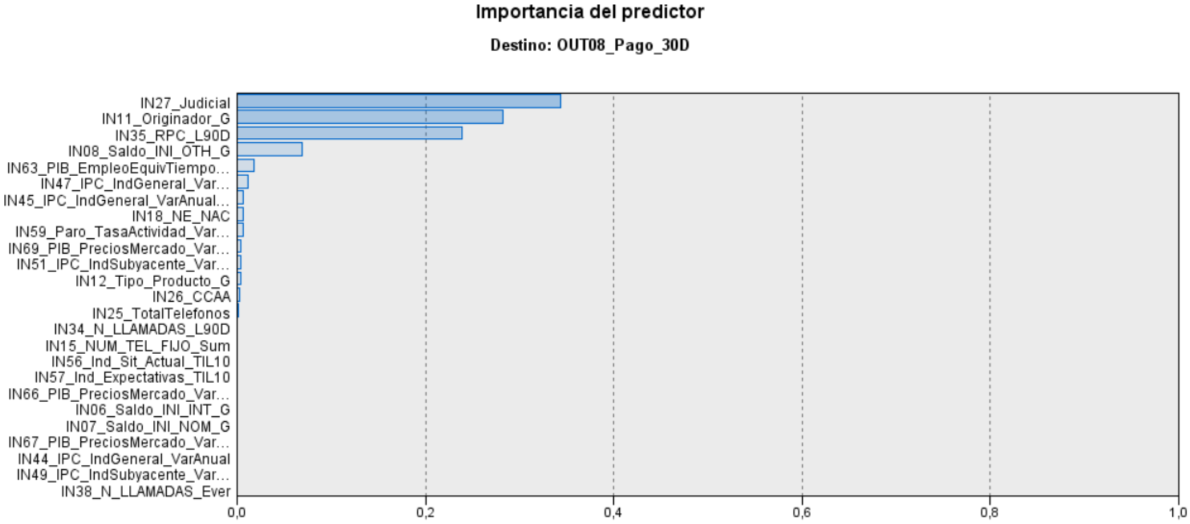
Table 250. Variables included in the Individuals Global model (complete)

Variable	Description
"IN27_Judicial"	Judicial flag
"IN11_Originador_G"	Creditor grouped
"IN35_RPC_L90D"	Right-party contact in the last 90 days flag
"IN08_Saldo_INI_OTH_G"	Initial other debt amount
"IN63_PIB_EmpleoEquivTiempoCompleto_VarAnual_TIL10"	GDP Full-Time Equivalent Employment on annual variation grouped in deciles
"IN47_IPC_IndGeneral_VarMensual_TIL10"	General Consumer Price Index on monthly variation grouped in deciles
"IN45_IPC_IndGeneral_VarAnual_TIL10"	General Consumer Price Index on annual variation grouped in deciles
"IN18_NE_NAC"	National nationality flag
"IN59_Paro_TasaActividad_VarAnual_TIL10"	Activity Rate on annual variation grouped in deciles
"IN69_PIB_PreciosMercado_VarTrim_TIL10"	GDP Market Prices on quarterly variation grouped in deciles
"IN51_IPC_IndSubyacente_VarMensual_TIL10"	Underlying Consumer Price Index on monthly variation grouped in deciles
"IN12_Tipo_Producto_G"	Product type grouped
"IN26_CCAA"	Region
"IN25_TotalTelefonos"	Number of phone numbers
"IN34_N_LLAMADAS_L90D"	Calls in the last 90 days flag
"IN15_NUM_TEL_FIJO_Sum"	Number of fixed line phone numbers
"IN56_Ind_Sit_Actual_TIL10"	Consumer Confidence Index actual situation grouped in deciles
"IN57_Ind_Expectativas_TIL10"	Consumer Confidence Index expectations grouped in deciles
"IN66_PIB_PreciosMercado_VarAnual"	GDP Market Prices on annual variation
"IN06_Saldo_INI_INT_G"	Initial interest debt amount
"IN07_Saldo_INI_NOM_G"	Initial principal debt amount

“IN67_PIB_PreciosMercado_VarAnual_TIL10”	GDP Market Prices on annual variation grouped in deciles
“IN44_IPC_IndGeneral_VarAnual”	General Consumer Price Index on annual variation
“IN49_IPC_IndSubyacente_VarAnual_TIL10”	Underlying Consumer Price Index on annual variation grouped in deciles
“IN38_N_LLAMADAS_Ever”	Calls ever flag

Source: Own elaboration

Figure 250. Variables importance for the Individuals Global model (complete)

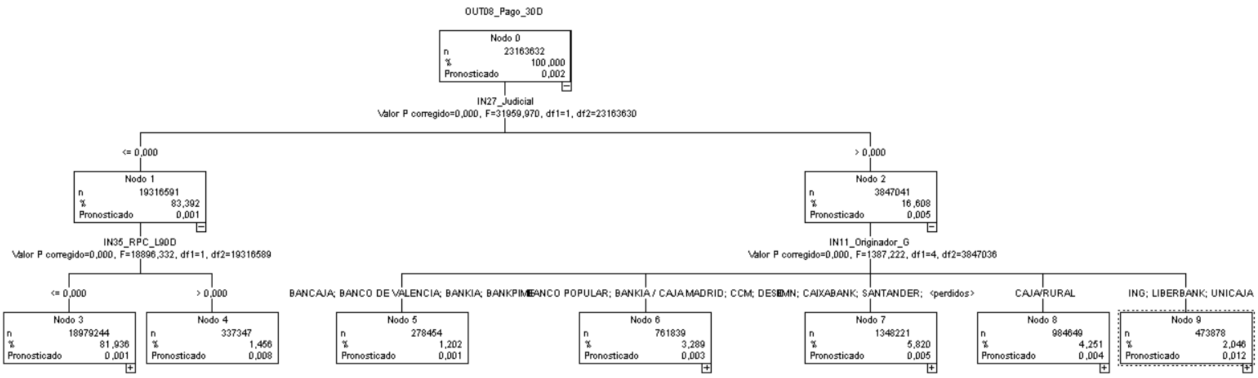


Source: Own elaboration

Model snapshot – initial layers

In the following figure, the initial 3 layers of this model are shown.

Figure 251. Initial layers of the Individuals Global model (complete)



Source: Own elaboration

Gini and KS indicators and model predictability

As anticipated in Section 4.3 of this document, we assess the predictability power and robustness of the models built by using the Gini indicator (Gini) and the Kolmogorov-Smirnoff test (KS). In the following table we summarize the Gini and KS indicators for the Individuals Global model.

Table 251. Gini and K-S indicators of the Individuals Global model (complete)

Indicators	Training sample	Testing sample
Gini	0,635	0,638
K-S	0,485	0,489

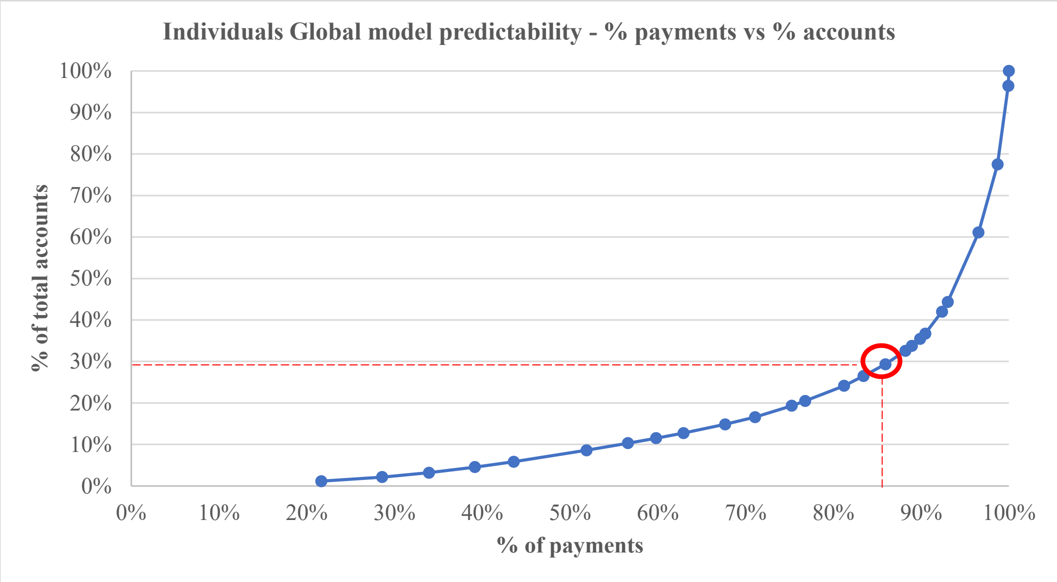
Source: Own elaboration

Model predictability: % of accounts and % of payments

From the common practice of the Collections and Recoveries industry, we applied another assessment of the predictability of the model: what percentage of positives cases of the variable to be explained (make a payment in the following 30 days) is the model able to identify with a small selection of cases.

Based on the model built for the Individuals Global, with a selection of 29% of the accounts, we obtain 86% of the total accounts which will pay in the following 30 days.

Figure 252. Distribution of accounts and payments on the Individuals Global model (complete)



Source: Own elaboration

Table 252. Distribution of accounts and payments on the Individuals Global model (complete)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	100%	96%
0,002	99%	77%
0,003	97%	61%
0,004	93%	44%
0,005	92%	42%
0,007	90%	37%
0,008	90%	35%
0,009	89%	34%
0,010	88%	33%
0,012	86%	29%
0,013	83%	26%
0,017	81%	24%
0,019	77%	20%
0,021	75%	19%
0,028	71%	17%
0,031	68%	15%
0,036	63%	13%
0,037	60%	12%
0,038	57%	10%
0,042	52%	9%
0,047	44%	6%
0,054	39%	5%
0,072	34%	3%
0,096	29%	2%
0,265	22%	1%

Source: Own elaboration

5.1.2.- Individuals Pre-Covid model (complete)

Variables included in the model and their importance as predictor

The variables included in this model are the following 26 (ordered in descendent importance) in Table 253.

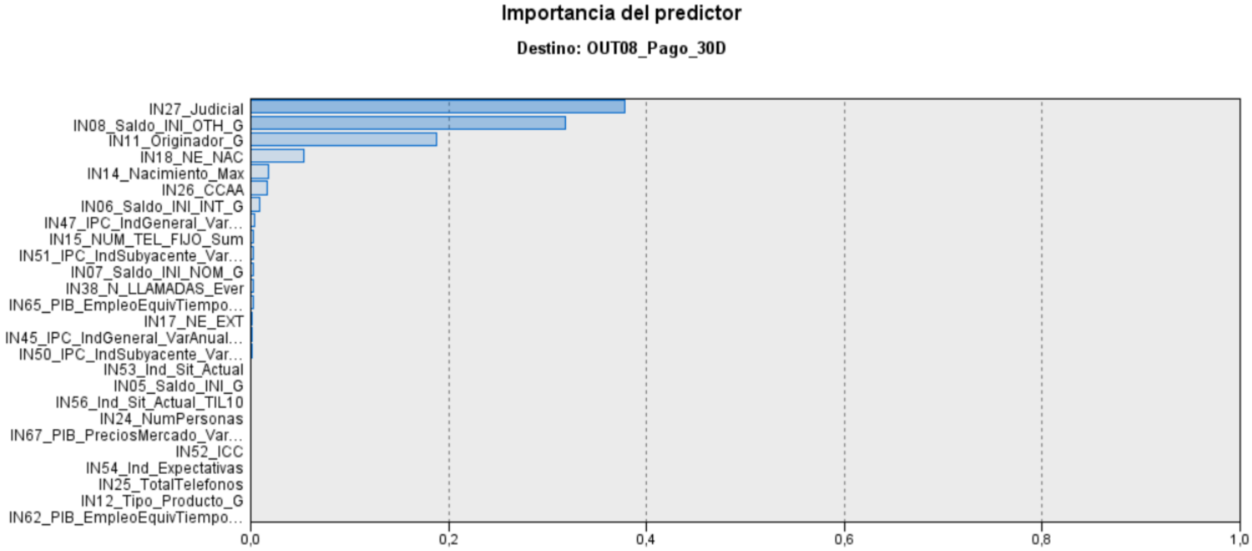
Table 253. Variables included in the Individuals Pre-Covid model (complete)

Variable	Description
"IN27_Judicial"	Judicial flag
"IN08_Saldo_INI_OTH_G"	Initial other debt amount
"IN11_Originador_G"	Creditor grouped
"IN18_NE_NAC"	National nationality flag
"IN14_Nacimiento_Max"	Youngest date of birth or date of company creation
"IN26_CCAA"	Region
"IN06_Saldo_INI_INT_G"	Initial interest debt amount
"IN47_IPC_IndGeneral_VarMensual_TIL10"	General Consumer Price Index on monthly variation grouped in deciles
"IN51_IPC_IndSubyacente_VarMensual_TIL10"	Underlying Consumer Price Index on monthly variation grouped in deciles
"IN07_Saldo_INI_NOM_G"	Initial principal debt amount
"IN38_N_LLAMADAS_Ever"	Calls ever flag
"IN65_PIB_EmpleoEquivTiempoCompleto_VarTrim_TIL10"	GDP Full-Time Equivalent Employment on quarterly variation grouped in deciles
"IN17_NE_EXT"	Foreign nationality flag
"IN45_IPC_IndGeneral_VarAnual_TIL10"	General Consumer Price Index on annual variation grouped in deciles
"IN50_IPC_IndSubyacente_VarMensual"	Underlying Consumer Price Index on monthly variation
"IN53_Ind_Sit_Actual"	Consumer Confidence Index actual situation
"IN05_Saldo_INI_G"	Initial total debt amount
"IN56_Ind_Sit_Actual_TIL10"	Consumer Confidence Index actual situation grouped in deciles
"IN24_NumPersonas"	Number of participants
"IN67_PIB_PreciosMercado_VarAnual_TIL10"	GDP Market Prices on annual variation grouped in deciles
"IN52_ICC"	Consumer Confidence Index
"IN54_Ind_Expectativas"	Consumer Confidence Index expectations
"IN25_TotalTelefonos"	Number of phone numbers
"IN12_Tipo_Producto_G"	Product type grouped

“IN62_PIB_EmpleoEquivTiempoCompleto_VarAnual”	GDP Full-Time Equivalent Employment on annual variation

Source: Own elaboration

Figure 253. Variables importance for the Individuals Pre-Covid model (complete)

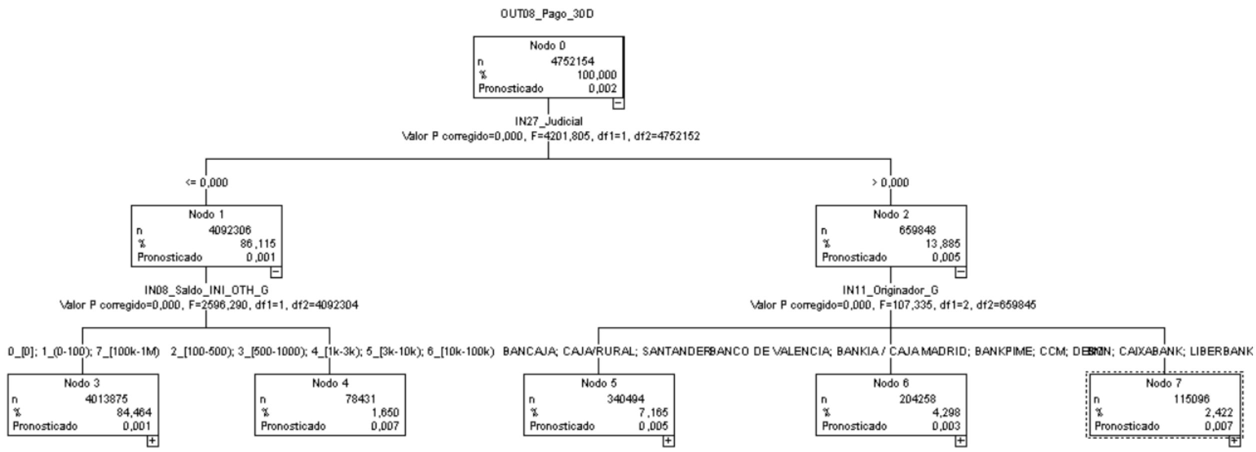


Source: Own elaboration

Model snapshot – initial layers

In the following figure, the initial 3 layers of this model are shown.

Figure 254. Initial layers of the Individuals Pre-Covid model (complete)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table we summarize the Gini and KS indicators for the Individuals Pre-Covid model.

Table 254. Gini and K-S indicators of the Individuals Pre-Covid model (complete)

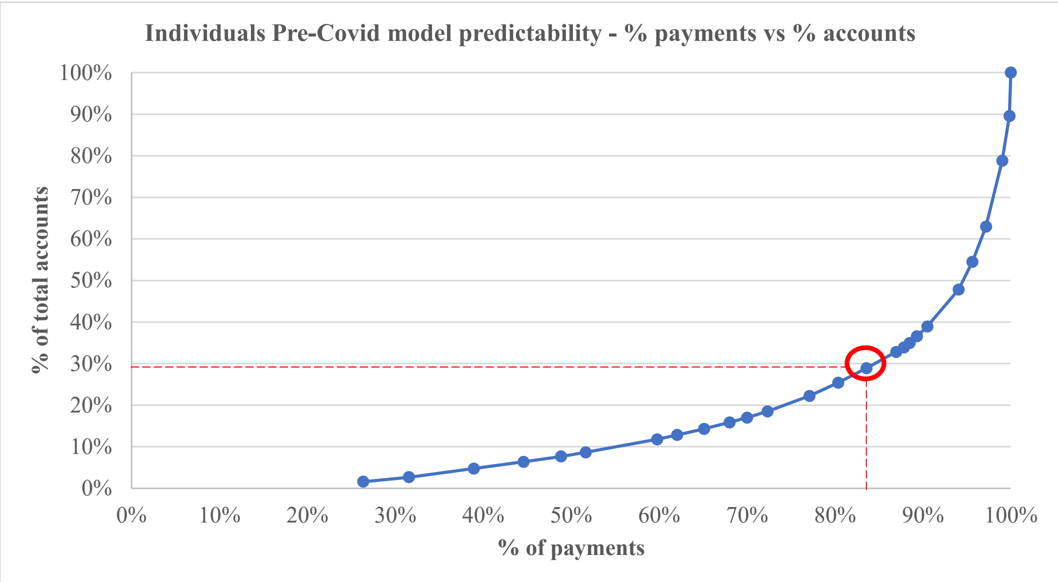
Indicators	Training sample	Testing sample
Gini	0,571	0,562
K-S	0,427	0,423

Source: Own elaboration

Model predictability: % of accounts and % of payments

Based on the model built for the Individuals Pre-Covid, with a selection of 29% of the accounts, we obtain 84% of the total accounts which will pay in the following 30 days.

Figure 255. Distribution of accounts and payments on the Individuals Pre-Covid model (complete)



Source: Own elaboration

Table 255. Distribution of accounts and payments on the Individuals Pre-Covid model (complete)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	100%	90%
0,002	99%	79%
0,003	97%	63%
0,004	96%	54%
0,006	94%	48%
0,007	90%	39%
0,008	89%	37%
0,009	88%	35%
0,012	88%	34%
0,013	87%	33%
0,014	84%	29%
0,016	80%	25%
0,020	77%	22%
0,024	72%	19%
0,027	70%	17%
0,028	68%	16%
0,032	65%	14%
0,033	62%	13%
0,040	60%	12%
0,043	52%	9%
0,051	49%	8%
0,053	45%	6%
0,054	39%	5%
0,074	32%	3%
0,252	26%	2%

Source: Own elaboration

5.1.3.- Individuals Covid model (complete)

Variables included in the model and their importance as predictor

The variables included in this model are the following 24 (ordered in descendent importance) in Table 256.

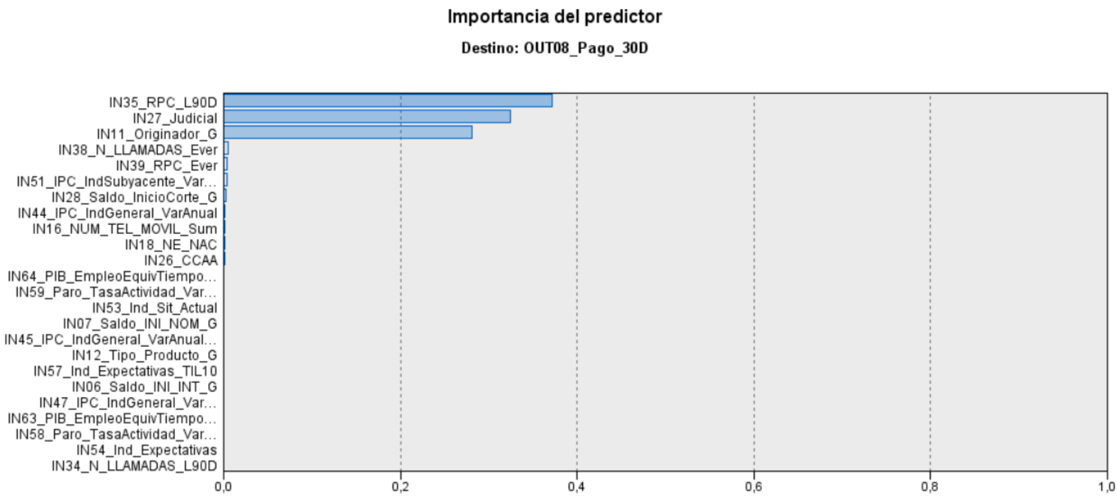
Table 256. Variables included in the Individuals Covid model (complete)

Variable	Description
"IN35_RPC_L90D"	Right-party contact in the last 90 days flag
"IN27_Judicial"	Judicial flag
"IN11_Originador_G"	Creditor grouped
"IN38_N_LLAMADAS_Ever"	Calls ever flag
"IN39_RPC_Ever"	Right-party contact ever flag

“IN51_IPC_IndSubyacente_VarMensual_TIL10”	Underlying Consumer Price Index on monthly variation grouped in deciles
“IN28_Saldo_InicioCorte_G”	Monthly debt amount grouped
“IN44_IPC_IndGeneral_VarAnual”	General Consumer Price Index on annual variation
“IN16_NUM_TEL_MOVIL_Sum”	Number of mobile phone numbers
“IN18_NE_NAC”	National nationality flag
“IN26_CCAA”	Region
“IN64_PIB_EmpleoEquivTiempoCompleto_VarTrim”	GDP Full-Time Equivalent Employment on quarterly variation
“IN59_Paro_TasaActividad_VarAnual_TIL10”	Activity Rate on annual variation grouped in deciles
“IN53_Ind_Sit_Actual”	Consumer Confidence Index actual situation
“IN07_Saldo_INI_NOM_G”	Initial principal debt amount
“IN45_IPC_IndGeneral_VarAnual_TIL10”	General Consumer Price Index on annual variation grouped in deciles
“IN12_Tipo_Producto_G”	Product type grouped
“IN57_Ind_Expectativas_TIL10”	Consumer Confidence Index expectations grouped in deciles
“IN06_Saldo_INI_INT_G”	Initial interest debt amount
“IN47_IPC_IndGeneral_VarMensual_TIL10”	General Consumer Price Index on monthly variation grouped in deciles
“IN63_PIB_EmpleoEquivTiempoCompleto_VarAnual_TIL10”	GDP Full-Time Equivalent Employment on annual variation grouped in deciles
“IN58_Paro_TasaActividad_VarAnual”	Activity Rate on annual variation
“IN54_Ind_Expectativas”	Consumer Confidence Index expectations
“IN34_N_LLAMADAS_L90D”	Calls in the last 90 days flag

Source: Own elaboration

Figure 256. Variables importance for the Individuals Covid model (complete)

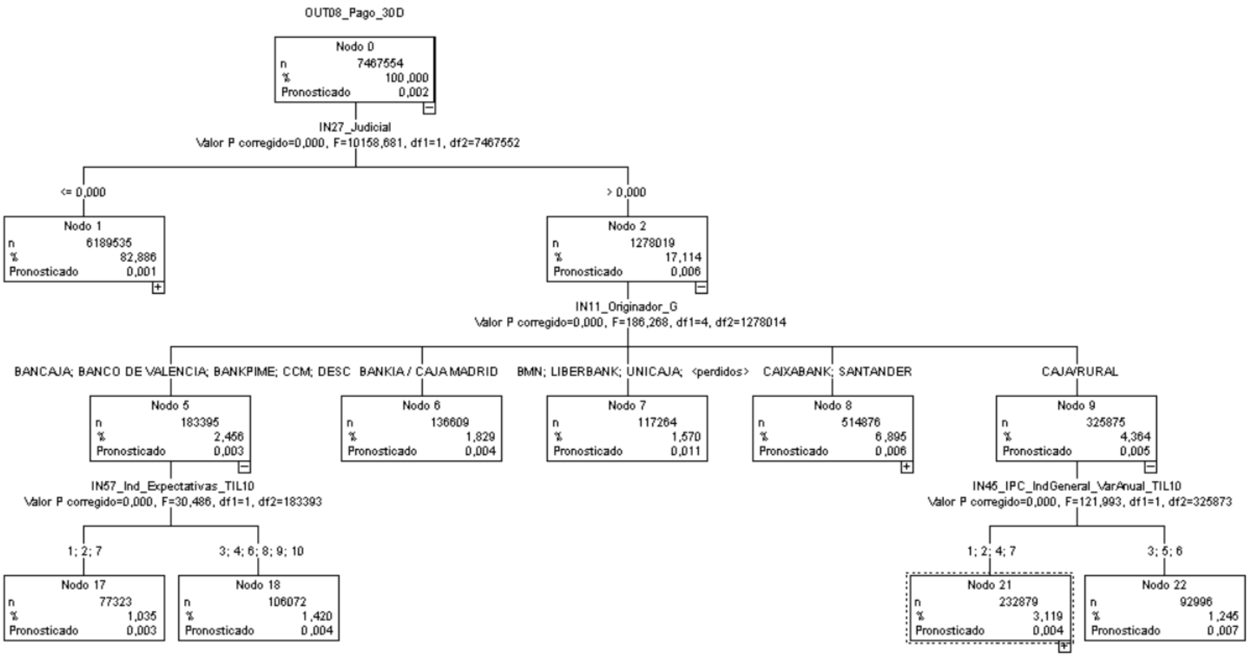


Source: Own elaboration

Model snapshot – initial layers

In the following figure, the initial 3 layers of this model are shown.

Figure 257. Initial layers of the Individuals Covid model (complete)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table we summarize the Gini and KS indicators for the Individuals Covid model.

Table 257. Gini and K-S indicators of the Individuals Covid model (complete)

Indicators	Training sample	Testing sample
Gini	0,611	0,611
K-S	0,467	0,477

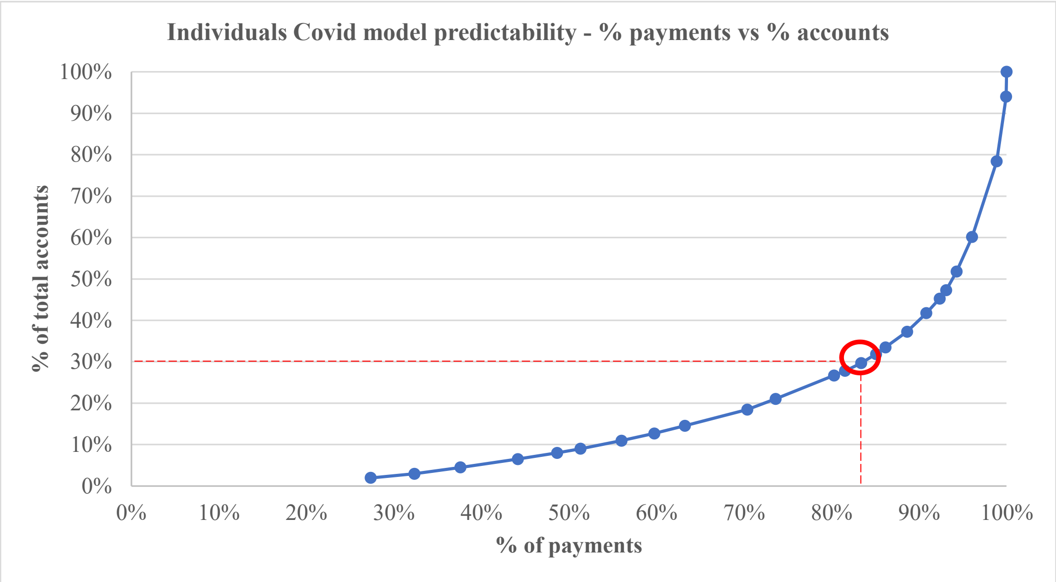
Source: Own elaboration

Model predictability: % of accounts and % of payments

Based on the model built for the Individuals Covid, with a selection of 30% of the accounts, we obtain 83% of the total accounts which will pay in the following 30 days.

Table 258 and Figure 258. Distribution of accounts and payments on the Individuals Covid model (complete)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	100%	94%
0,002	99%	78%
0,003	96%	60%
0,004	94%	52%
0,005	93%	47%
0,006	92%	45%
0,007	91%	42%
0,009	89%	37%
0,010	86%	33%
0,011	85%	32%
0,014	83%	30%
0,016	82%	28%
0,017	80%	27%
0,018	74%	21%
0,026	70%	18%
0,027	63%	15%
0,031	60%	13%
0,034	56%	11%
0,038	51%	9%
0,043	49%	8%
0,047	44%	6%
0,049	38%	4%
0,070	32%	3%
0,200	27%	2%



Source: Own elaboration

5.1.4.- Individuals Post-Covid model (complete)

Variables included in the model and their importance as predictor

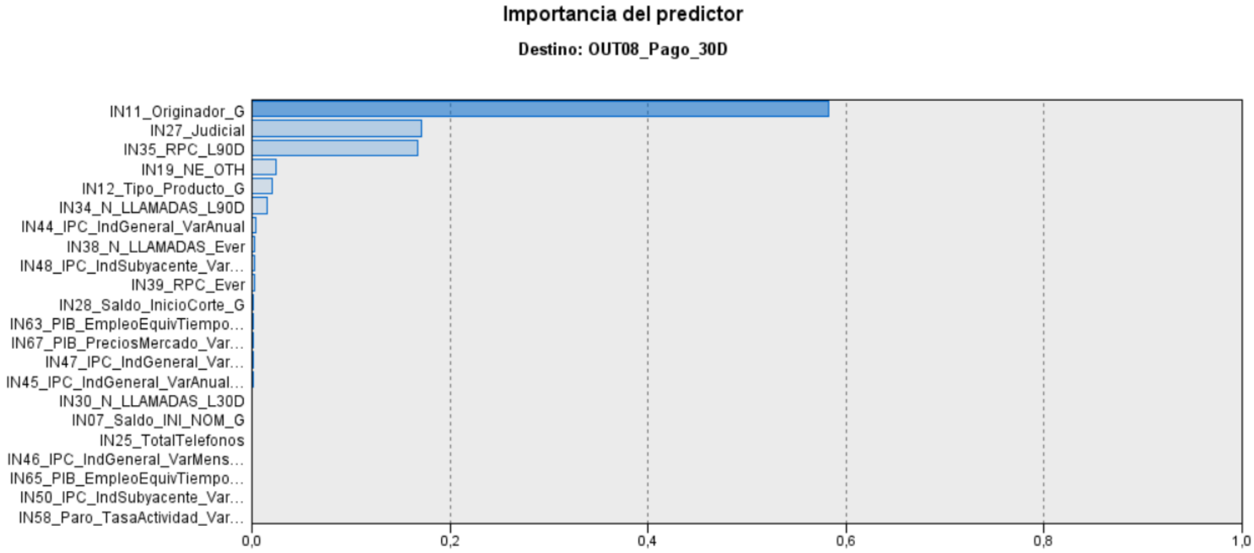
The variables included in this model are the following 22 (ordered in descendent importance) in Table 259.

Table 259. Variables included in the Individuals Post-Covid model (complete)

Variable	Description
"IN11_Originador_G"	Creditor grouped
"IN27_Judicial"	Judicial flag
"IN35_RPC_L90D"	Right-party contact in the last 90 days flag
"IN19_NE_OTH"	Other nationality flag
"IN12_Tipo_Producto_G"	Product type grouped
"IN34_N_LLAMADAS_L90D"	Calls in the last 90 days flag
"IN44_IPC_IndGeneral_VarAnual"	General Consumer Price Index on annual variation
"IN38_N_LLAMADAS_Ever"	Calls ever flag
"IN48_IPC_IndSubyacente_VarAnual"	Underlying Consumer Price Index on annual variation
"IN39_RPC_Ever"	Right-party contact ever flag
"IN28_Saldo_InicioCorte_G"	Monthly debt amount grouped
"IN63_PIB_EmpleoEquivTiempoCompleto_VarAnual_TIL10"	GDP Full-Time Equivalent Employment on annual variation grouped in deciles
"IN67_PIB_PreciosMercado_VarAnual_TIL10"	GDP Market Prices on annual variation grouped in deciles
"IN47_IPC_IndGeneral_VarMensual_TIL10"	General Consumer Price Index on monthly variation grouped in deciles
"IN45_IPC_IndGeneral_VarAnual_TIL10"	General Consumer Price Index on annual variation grouped in deciles
"IN30_N_LLAMADAS_L30D"	Calls in the last 30 days flag
"IN07_Saldo_INI_NOM_G"	Initial principal debt amount
"IN25_TotalTelefonos"	Number of phone numbers
"IN46_IPC_IndGeneral_VarMensual"	General Consumer Price Index on monthly variation
"IN65_PIB_EmpleoEquivTiempoCompleto_VarTrim_TIL10"	GDP Full-Time Equivalent Employment on quarterly variation grouped in deciles
"IN50_IPC_IndSubyacente_VarMensual"	Underlying Consumer Price Index on monthly variation
"IN58_Paro_TasaActividad_VarAnual"	Activity Rate on annual variation

Source: Own elaboration

Figure 259. Variables importance for the Individuals Post-Covid model (complete)

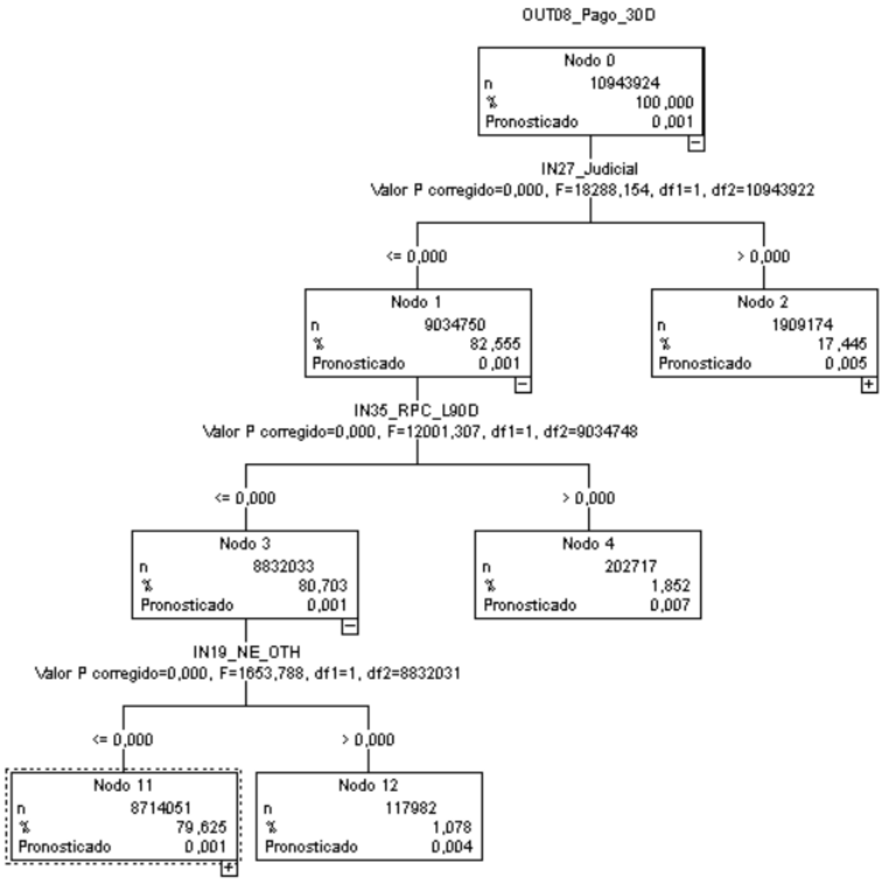


Source: Own elaboration

Model snapshot – initial layers

In the following figure, the initial 3 layers of this model are shown.

Figure 260. Initial layers of the Individuals Post-Covid model (complete)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table we summarize the Gini and KS indicators for the Individuals Post-Covid model.

Table 260. Gini and K-S indicators of the Individuals Post-Covid model (complete)

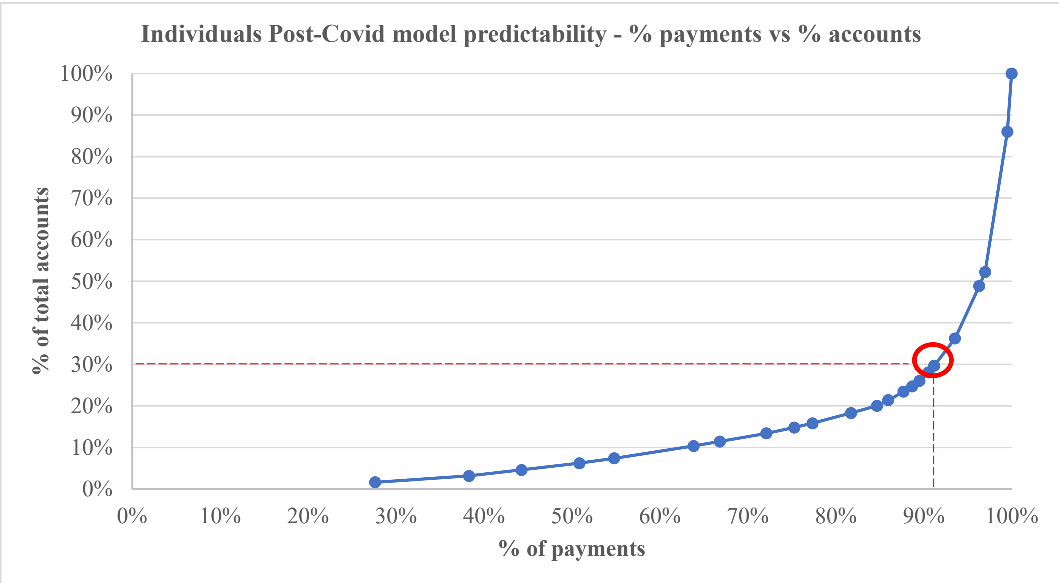
Indicators	Training sample	Testing sample
Gini	0,714	0,714
K-S	0,573	0,577

Source: Own elaboration

Model predictability: % of accounts and % of payments

Based on the model built for the Individuals Post-Covid, with a selection of 30% of the accounts, we obtain 91% of the total accounts which will pay in the following 30 days.

Figure 261. Distribution of accounts and payments on the Individuals Pre-Covid model (complete)



Source: Own elaboration

Table 261. Distribution of accounts and payments on the Individuals Pre-Covid model (complete)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	100%	86%
0,002	97%	52%
0,003	96%	49%
0,004	94%	36%
0,005	91%	30%
0,006	91%	28%
0,007	90%	26%
0,009	89%	25%
0,010	88%	23%
0,011	86%	21%
0,020	85%	20%
0,021	82%	18%
0,024	77%	16%
0,026	75%	15%
0,031	72%	13%
0,033	67%	11%
0,035	64%	10%
0,040	55%	7%
0,047	51%	6%
0,049	44%	5%
0,081	38%	3%
0,199	28%	2%

Source: Own elaboration

5.1.5.- SMEs Global model (complete)

Variables included in the model and their importance as predictor

The variables included in this model are the following 21 (ordered in descendent importance) in Table 262.

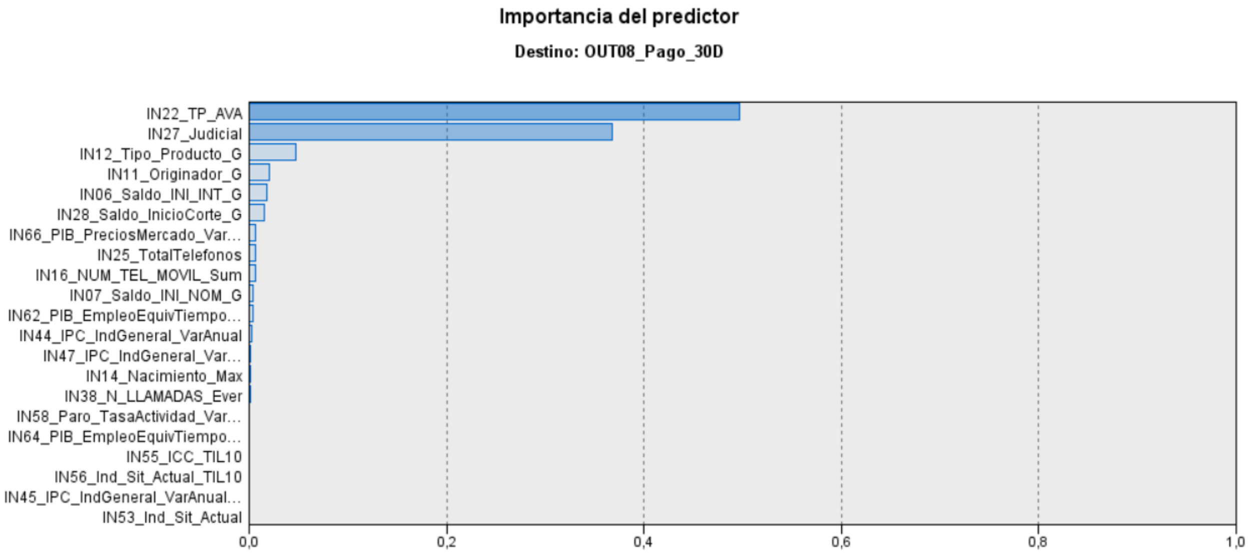
Table 262. Variables included in the SMEs Global model (complete)

Variable	Description
“IN22_TP_AVA”	Guarantor flag
“IN27_Judicial”	Judicial flag
“IN12_Tipo_Producto_G”	Product type grouped
“IN11_Originador_G”	Creditor grouped
“IN06_Saldo_INI_INT_G”	Initial interest debt amount
“IN28_Saldo_InicioCorte_G”	Monthly debt amount grouped
“IN66_PIB_PreciosMercado_VarAnual”	GDP Market Prices on annual variation

“IN25_TotalTelefonos”	Number of phone numbers
“IN16_NUM_TEL_MOVIL_Sum”	Number of mobile phone numbers
“IN07_Saldo_INI_NOM_G”	Initial principal debt amount
“IN62_PIB_EmpleoEquivTiempoCompleto_VarAnual”	GDP Full-Time Equivalent Employment on annual variation
“IN44_IPC_IndGeneral_VarAnual”	General Consumer Price Index on annual variation
“IN47_IPC_IndGeneral_VarMensual_TIL10”	General Consumer Price Index on monthly variation grouped in deciles
“IN14_Nacimiento_Max”	Youngest date of birth or date of company creation of participants
“IN38_N_LLAMADAS_Ever”	Calls ever flag
“IN58_Paro_TasaActividad_VarAnual”	Activity Rate on annual variation
“IN64_PIB_EmpleoEquivTiempoCompleto_VarTrim”	GDP Full-Time Equivalent Employment on quarterly variation
“IN55_ICC_TIL10”	Consumer Confidence Index grouped in deciles
“IN56_Ind_Sit_Actual_TIL10”	Consumer Confidence Index actual situation grouped in deciles
“IN45_IPC_IndGeneral_VarAnual_TIL10”	General Consumer Price Index on annual variation grouped in deciles
“IN53_Ind_Sit_Actual”	Consumer Confidence Index actual situation

Source: Own elaboration

Figure 262. Variables importance for the SMEs Global model (complete)

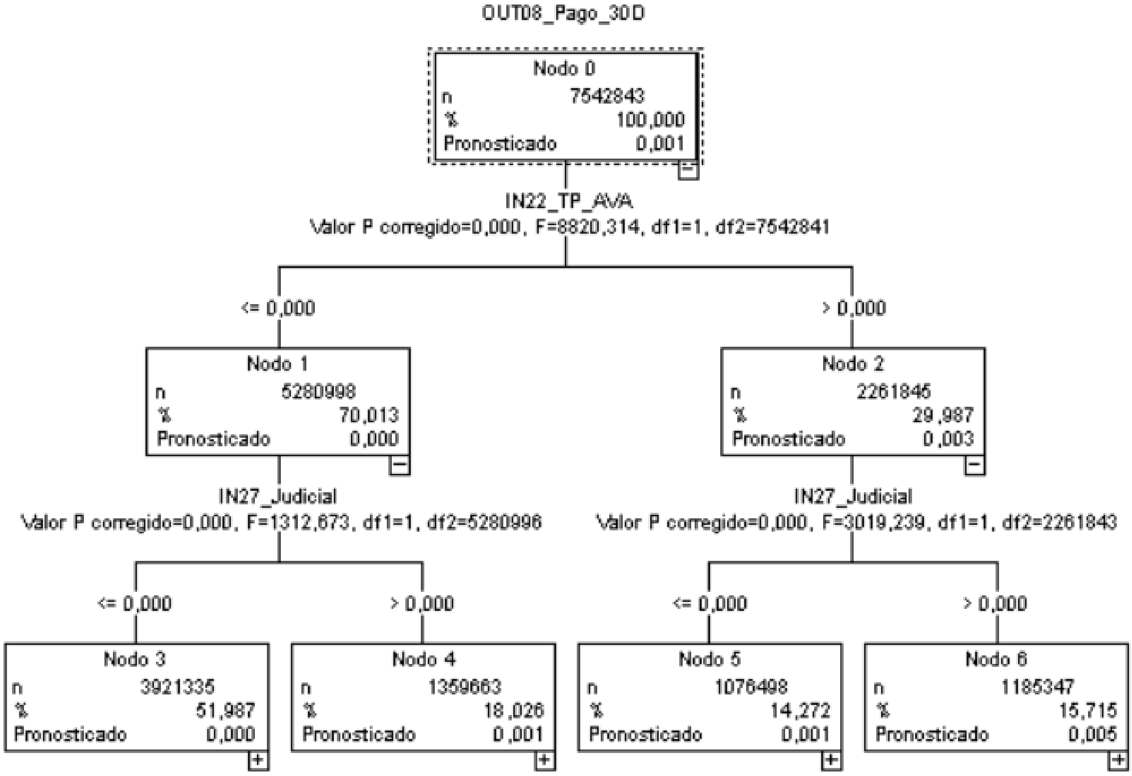


Source: Own elaboration

Model snapshot – initial layers

In the following figure, the initial 3 layers of this model are shown.

Figure 263. Initial layers of the SMEs Global model (complete)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table we summarize the Gini and KS indicators for the SMEs Global model.

Table 263. Gini and K-S indicators of the SMEs Global model (complete)

Indicators	Training sample	Testing sample
Gini	0,662	0,639
K-S	0,529	0,515

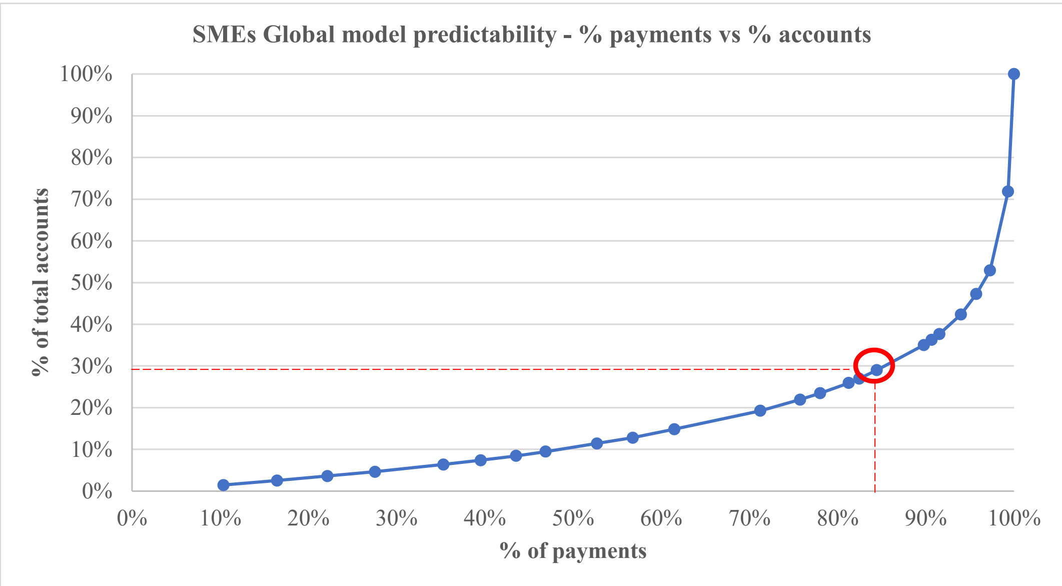
Source: Own elaboration

Model predictability: % of accounts and % of payments

Based on the model built for the SMEs Global, with a selection of 29% of the accounts, we obtain 84% of the total accounts which will pay in the following 30 days.

Table 264 and Figure 264. Distribution of accounts and payments on the SMEs Global model (complete)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	99%	72%
0,002	97%	53%
0,003	96%	47%
0,004	94%	42%
0,005	92%	38%
0,006	91%	36%
0,007	90%	35%
0,008	84%	29%
0,009	82%	27%
0,011	81%	26%
0,012	78%	23%
0,014	76%	22%
0,018	71%	19%
0,019	61%	15%
0,024	57%	13%
0,025	53%	11%
0,027	47%	9%
0,031	44%	8%
0,035	40%	7%
0,037	35%	6%
0,043	28%	5%
0,044	22%	4%
0,046	16%	3%
0,060	10%	1%



Source: Own elaboration

5.1.6.- SMEs Pre-Covid model (complete)

Variables included in the model and their importance as predictor

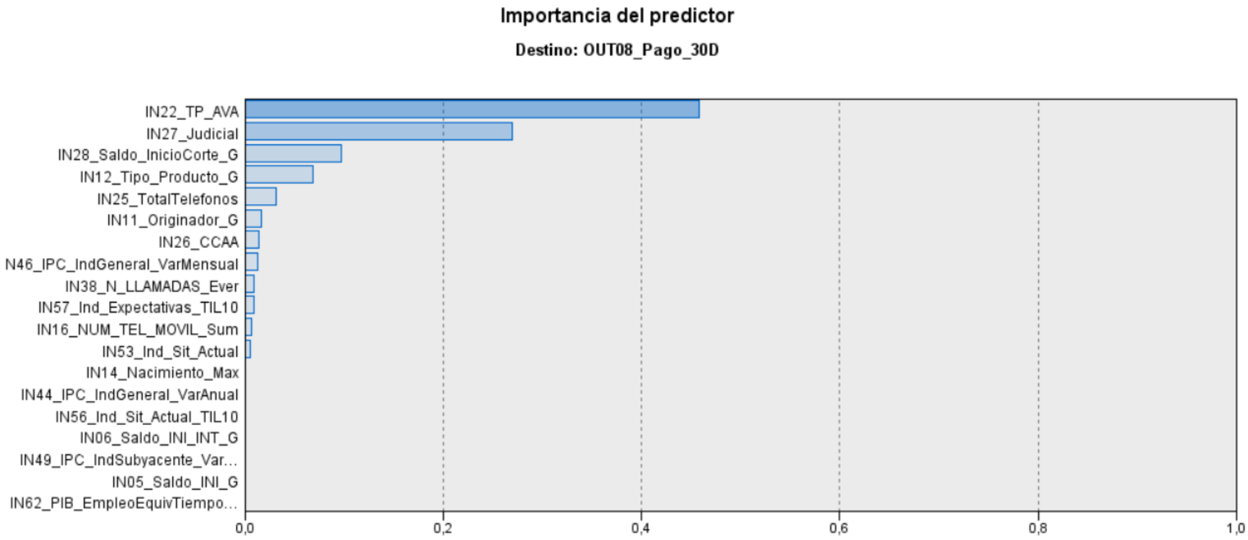
The variables included in this model are the following 19 (ordered in descendent importance) in Table 265.

Table 265. Variables included in the SMEs Pre-Covid model (complete)

Variable	Description
"IN22_TP_AVA"	Guarantor flag
"IN27_Judicial"	Judicial flag
"IN28_Saldo_InicioCorte_G"	Monthly debt amount grouped
"IN12_Tipo_Producto_G"	Product type grouped
"IN25_TotalTelefonos"	Number of phone numbers
"IN11_Originador_G"	Creditor grouped
"IN26_CCAA"	Region
"IN46_IPC_IndGeneral_VarMensual"	General Consumer Price Index on monthly variation
"IN38_N_LLAMADAS_Ever"	Calls ever flag
"IN57_Ind_Expectativas_TIL10"	Consumer Confidence Index expectations grouped in deciles
"IN16_NUM_TEL_MOVIL_Sum"	Number of mobile phone numbers
"IN53_Ind_Sit_Actual"	Consumer Confidence Index actual situation
"IN14_Nacimiento_Max"	Youngest date of birth or date of company creation of participants
"IN44_IPC_IndGeneral_VarAnual"	General Consumer Price Index on annual variation
"IN56_Ind_Sit_Actual_TIL10"	Consumer Confidence Index actual situation grouped in deciles
"IN06_Saldo_INI_INT_G"	Initial interest debt amount
"IN49_IPC_IndSubyacente_VarAnual_TIL10"	Underlying Consumer Price Index on annual variation grouped in deciles
"IN05_Saldo_INI_G"	Initial total debt amount
"IN62_PIB_EmpleoEquivTiempoCompleto_VarAnual"	GDP Full-Time Equivalent Employment on annual variation

Source: Own elaboration

Figure 265. Variables importance for the SMEs Pre-Covid model (complete)

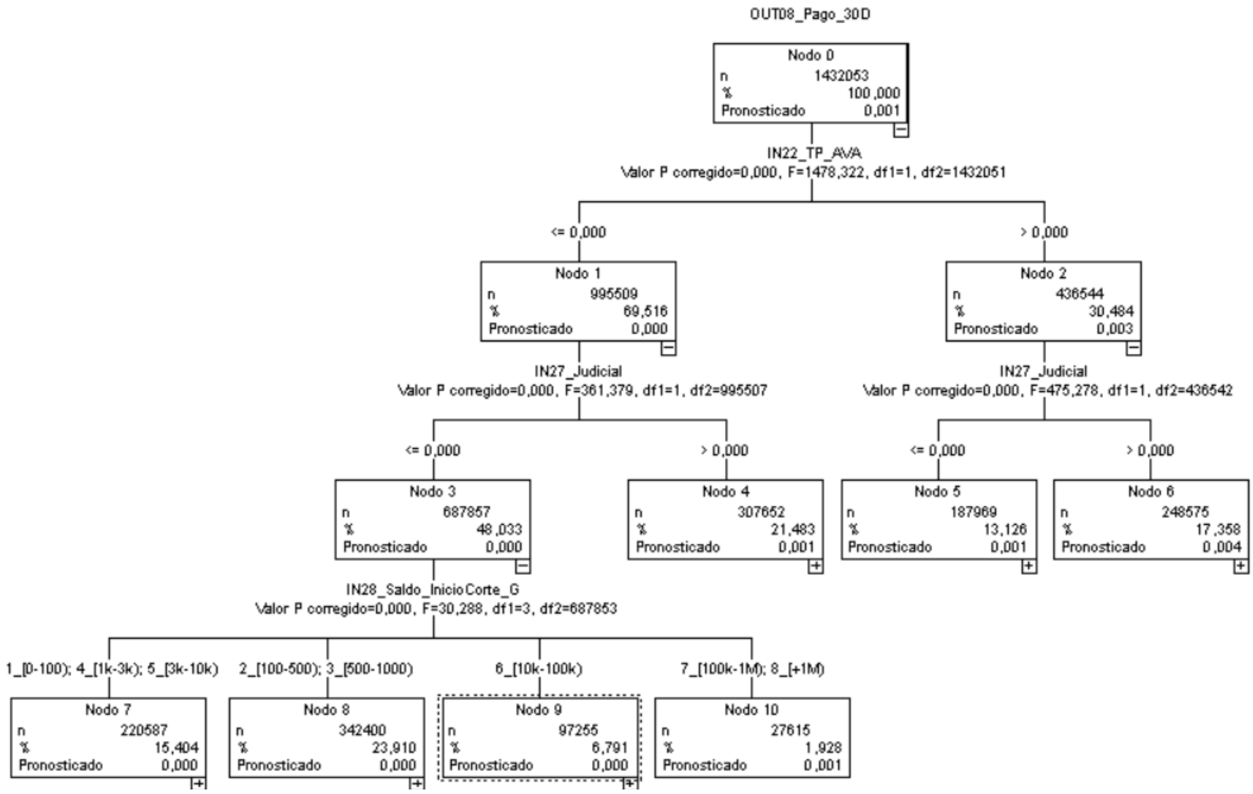


Source: Own elaboration

Model snapshot – initial layers

In the following figure, the initial 3 layers of this model are shown.

Figure 266. Initial layers of the SMEs Pre-Covid model (complete)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table we summarize the Gini and KS indicators for the SMEs Pre-Covid model.

Table 266. Gini and K-S indicators of the SMEs Pre-Covid model (complete)

Indicators	Training sample	Testing sample
Gini	0,674	0,610
K-S	0,532	0,492

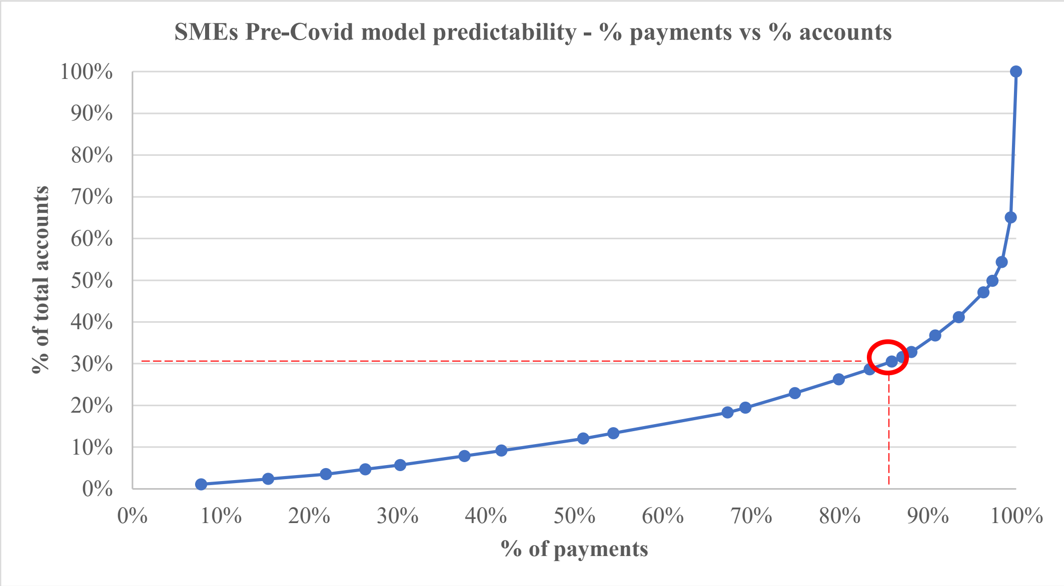
Source: Own elaboration

Model predictability: % of accounts and % of payments

Based on the model built for the SMEs Pre-Covid, with a selection of 31% of the accounts, we obtain 86% of the total accounts which will pay in the following 30 days.

Table 267 and Figure 267. Distribution of accounts and payments on the SMEs Pre-Covid model (complete)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	99%	65%
0,002	98%	54%
0,003	97%	50%
0,004	96%	47%
0,005	94%	41%
0,006	91%	37%
0,007	88%	33%
0,010	87%	32%
0,011	86%	31%
0,012	83%	29%
0,013	80%	26%
0,014	75%	23%
0,015	69%	19%
0,022	67%	18%
0,023	54%	13%
0,027	51%	12%
0,028	42%	9%
0,029	38%	8%
0,032	30%	6%
0,033	26%	5%
0,048	22%	4%
0,050	15%	2%
0,061	8%	1%



Source: Own elaboration

5.1.7.- SMEs Covid model (complete)

Variables included in the model and their importance as predictor

The variables included in this model are the following 24 (ordered in descendent importance) in Table 268.

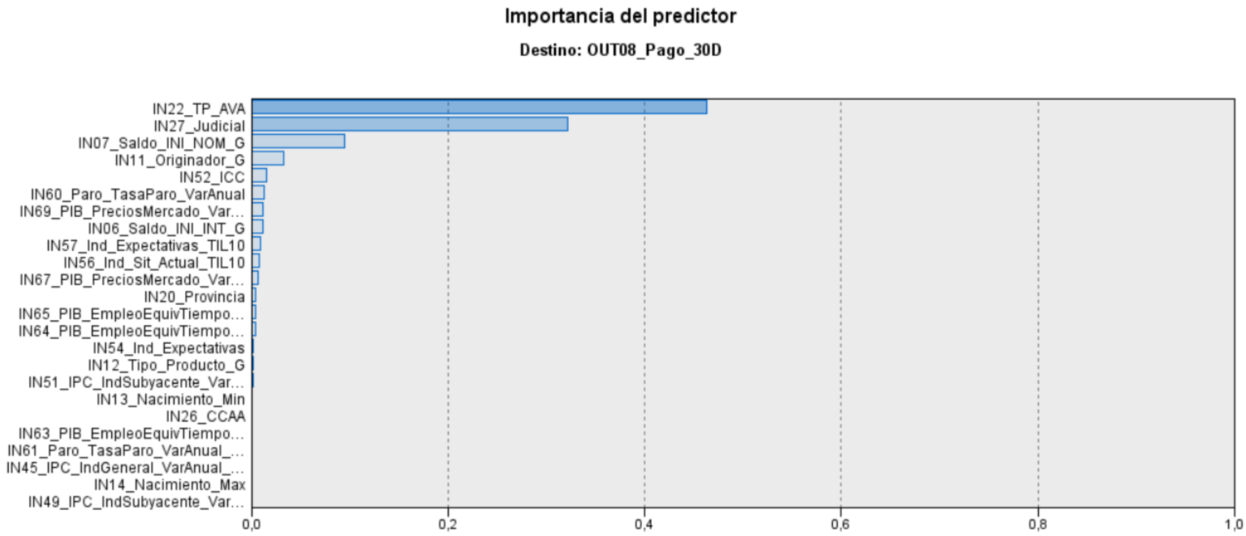
Table 268. Variables included in the SMEs Covid model (complete)

Variable	Description
"IN22_TP_AVA"	Guarantor flag
"IN27_Judicial"	Judicial flag
"IN07_Saldo_INI_NOM_G"	Initial principal debt amount
"IN11_Originador_G"	Creditor grouped
"IN52_ICC"	Consumer Confidence Index
"IN60_Paro_TasaParo_VarAnual"	Unemployment Rate on annual variation
"IN69_PIB_PreciosMercado_VarTrim_TIL10"	GDP Market Prices on quarterly variation grouped in deciles
"IN06_Saldo_INI_INT_G"	Initial interest debt amount
"IN57_Ind_Expectativas_TIL10"	Consumer Confidence Index expectations grouped in deciles
"IN56_Ind_Sit_Actual_TIL10"	Consumer Confidence Index actual situation grouped in deciles
"IN67_PIB_PreciosMercado_VarAnual_TIL10"	GDP Market Prices on annual variation grouped in deciles
"IN20_Provincia"	Province

"IN65_PIB_EmpleoEquivTiempoCompleto_VarTrim_TIL10"	GDP Full-Time Equivalent Employment on quarterly variation grouped in deciles
"IN64_PIB_EmpleoEquivTiempoCompleto_VarTrim"	GDP Full-Time Equivalent Employment on quarterly variation
"IN54_Ind_Expectativas"	Consumer Confidence Index expectations
"IN12_Tipo_Producto_G"	Product type grouped
"IN51_IPC_IndSubyacente_VarMensual_TIL10"	Underlying Consumer Price Index on monthly variation grouped in deciles
"IN13_Nacimiento_Min"	Oldest date of birth or date of company creation of participants
"IN26_CCAA"	Region
"IN63_PIB_EmpleoEquivTiempoCompleto_VarAnual_TIL10"	GDP Full-Time Equivalent Employment on annual variation grouped in deciles
"IN61_Paro_TasaParo_VarAnual_TIL10"	Unemployment Rate on annual variation grouped in deciles
"IN45_IPC_IndGeneral_VarAnual_TIL10"	General Consumer Price Index on annual variation grouped in deciles
"IN14_Nacimiento_Max"	Youngest date of birth or date of company creation of participants
"IN49_IPC_IndSubyacente_VarAnual_TIL10"	Underlying Consumer Price Index on annual variation grouped in deciles

Source: Own elaboration

Figure 268. Variables importance for the SMEs Covid model (complete)

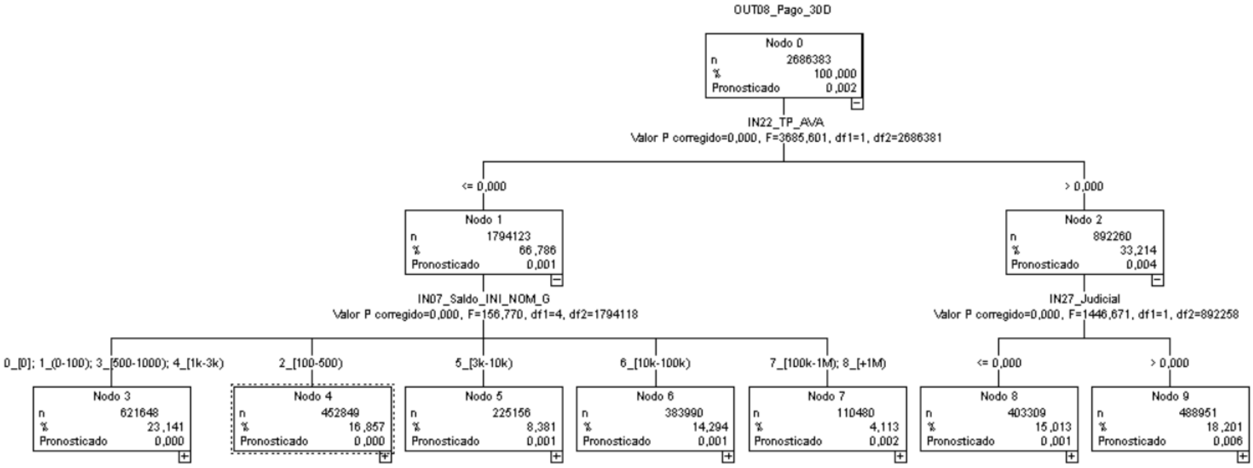


Source: Own elaboration

Model snapshot – initial layers

In the following figure, the initial 3 layers of this model are shown.

Figure 269. Initial layers of the SMEs Covid model (complete)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table we summarize the Gini and KS indicators for the SMEs Covid model.

Table 269. Gini and K-S indicators of the SMEs Covid model (complete)

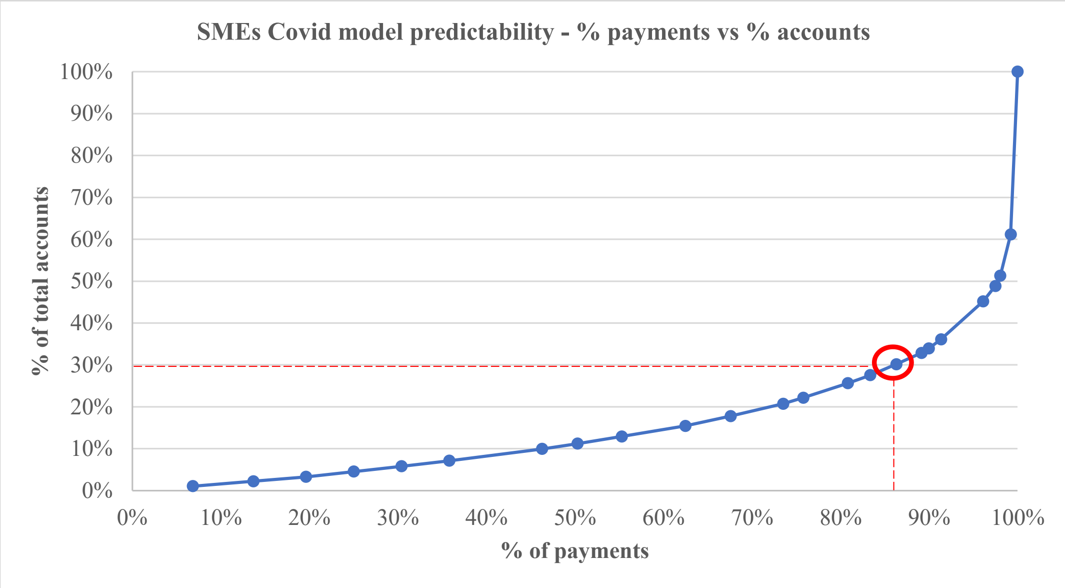
Indicators	Training sample	Testing sample
Gini	0,677	0,646
K-S	0,542	0,514

Source: Own elaboration

Model predictability: % of accounts and % of payments

Based on the model built for the SMEs Covid, with a selection of 30% of the accounts, we obtain 86% of the total accounts which will pay in the following 30 days.

Figure 270. Distribution of accounts and payments on the SMEs Covid model (complete)



Source: Own elaboration

Table 270. Distribution of accounts and payments on the SMEs Covid model (complete)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	99%	61%
0,002	98%	51%
0,003	97%	49%
0,004	96%	45%
0,005	91%	36%
0,006	90%	34%
0,008	89%	33%
0,009	86%	30%
0,010	83%	28%
0,011	81%	26%
0,012	76%	22%
0,016	74%	21%
0,017	68%	18%
0,022	62%	15%
0,023	55%	13%
0,025	50%	11%
0,029	46%	10%
0,032	36%	7%
0,033	30%	6%
0,034	25%	5%
0,044	20%	3%
0,046	14%	2%
0,051	7%	1%

Source: Own elaboration

5.1.8.- SMEs Post-Covid model (complete)

Variables included in the model and their importance as predictor

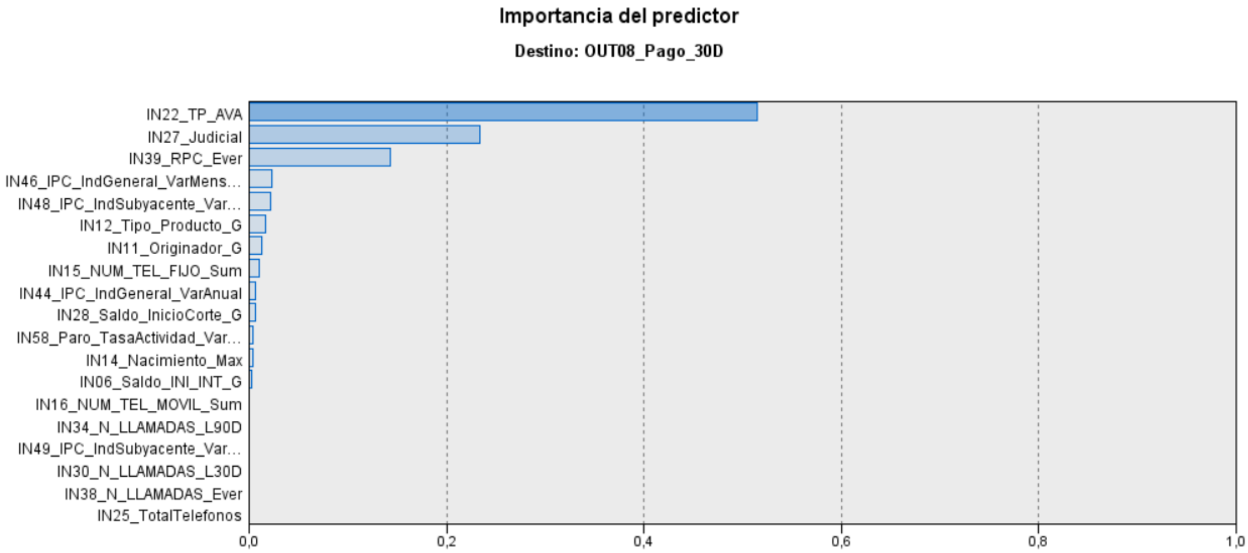
The variables included in this model are the following 19 (ordered in descendent importance) in Table 271.

Table 271. Variables included in the SMEs Post-Covid model (complete)

Variable	Description
"IN22_TP_AVA"	Guarantor flag
"IN27_Judicial"	Judicial flag
"IN39_RPC_Ever"	Right-party contact ever flag
"IN46_IPC_IndGeneral_VarMensual"	General Consumer Price Index on monthly variation
"IN48_IPC_IndSubyacente_VarAnual"	Underlying Consumer Price Index on annual variation
"IN12_Tipo_Producto_G"	Product type grouped
"IN11_Originador_G"	Creditor grouped
"IN15_NUM_TEL_FIJO_Sum"	Number of fixed line phone numbers
"IN44_IPC_IndGeneral_VarAnual"	General Consumer Price Index on annual variation
"IN28_Saldo_InicioCorte_G"	Monthly debt amount grouped
"IN58_Paro_TasaActividad_VarAnual"	Activity Rate on annual variation
"IN14_Nacimiento_Max"	Youngest date of birth or date of company creation of participants
"IN06_Saldo_INI_INT_G"	Initial interest debt amount
"IN16_NUM_TEL_MOVIL_Sum"	Number of mobile phone numbers
"IN34_N_LLAMADAS_L90D"	Calls in the last 90 days flag
"IN49_IPC_IndSubyacente_VarAnual_TIL10"	Underlying Consumer Price Index on annual variation grouped in deciles
"IN30_N_LLAMADAS_L30D"	Calls in the last 30 days flag
"IN38_N_LLAMADAS_Ever"	Calls ever flag
"IN25_TotalTelefonos"	Number of phone numbers

Source: Own elaboration

Figure 271. Variables importance for the SMEs Post-Covid model (complete)

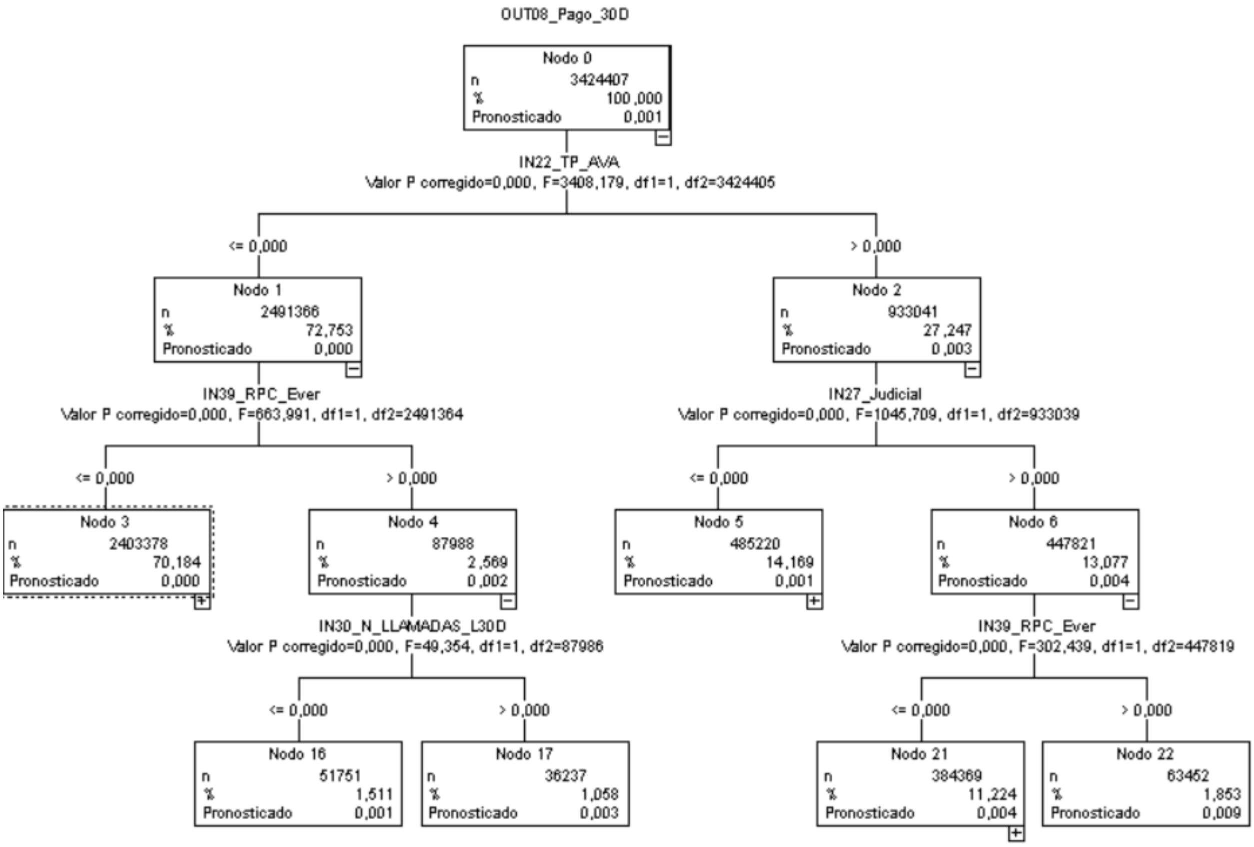


Source: Own elaboration

Model snapshot – initial layers

In the following figure, the initial 3 layers of this model are shown.

Figure 272. Initial layers of the SMEs Post-Covid model (complete)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table we summarize the Gini and KS indicators for the SMEs Post-Covid model.

Table 272. Gini and K-S indicators of the SMEs Post-Covid model (complete)

Indicators	Training sample	Testing sample
Gini	0,699	0,666
K-S	0,550	0,522

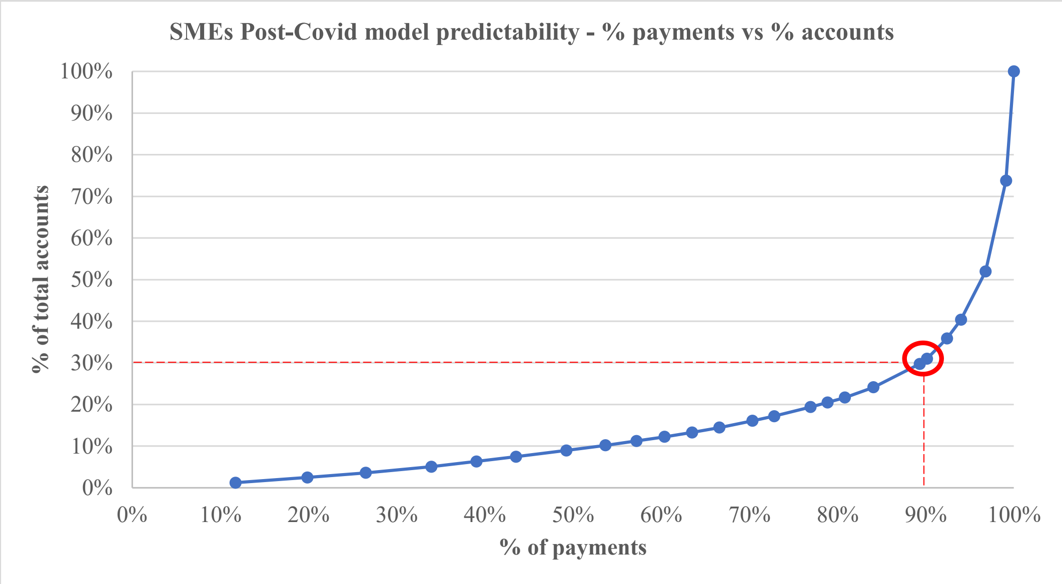
Source: Own elaboration

Model predictability: % of accounts and % of payments

Based on the model built for the SMEs Post-Covid, with a selection of 30% of the accounts, we obtain 89% of the total accounts which will pay in the following 30 days.

Table 273 and Figure 273. Distribution of accounts and payments on the SMEs Post-Covid model (complete)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	99%	74%
0,002	97%	52%
0,003	94%	40%
0,004	92%	36%
0,006	90%	31%
0,008	89%	30%
0,011	84%	24%
0,014	81%	22%
0,015	79%	20%
0,016	77%	19%
0,019	73%	17%
0,020	70%	16%
0,023	67%	14%
0,026	64%	13%
0,027	60%	12%
0,030	57%	11%
0,031	54%	10%
0,033	49%	9%
0,034	44%	7%
0,035	39%	6%
0,044	34%	5%
0,051	26%	4%
0,055	20%	2%
0,083	12%	1%



Source: Own elaboration

5.2.- Complete models with payers detail

5.2.1.- Summary of variables included in the models and their relevance (complete with payers)

In Table 274 we collect the summary of variables included in the Individuals complete with payers models, and these variables weight on the models; and in Table 275 we collect the summary of variables included in the SMEs complete with payers' models, and their weight on the models.

Table 274. Model variables and their weight for Individuals complete with payers models

Global		Pre-covid		Covid		Post-covid	
Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight
Payment l30d	0.9743	**Payment l30d**	0.9446	**Payment l30d**	0.9216	**Payment l90d**	0.9371
Judicial flag	0.0125	**Payment ever**	0.0403	**Payment ever**	0.0449	**Payment ever**	0.0221
Calls l90d flag	0.0058	**Judicial flag**	0.0037	**Judicial flag**	0.0248	**RPC l90d**	0.0135
Payment ever	0.0027	**Initial other debt amount grouped**	0.0018	**RPC l90d**	0.0048	**Creditor grouped**	0.0084
Initial debt amount grouped	0.0027	**Initial interest debt amount grouped**	0.0014	**RPC ever**	0.0027	**Judicial flag**	0.0075
GCPI annual	0.0007	**CCI**	0.0012	**Product type grouped**	0.0007	**Calls l90d flag**	0.0061
Creditor 99grouped	0.0007	**National nationality flag**	0.0011	**GDP MP quarter deciles**	0.0001	**Initial principal debt amount**	0.0029
GDP MP annual deciles	0.0002	**Product type grouped**	0.0010	**Monthly debt amount grouped**	0.0001	**RPC ever**	0.0020
Initial other debt amount grouped	0.0001	**Activity rate annual deciles**	0.0001	**CCI expectations**	0.0001	**Holder or active company flag**	0.0000
RPC l90d	0.0001	**GDP MP annual deciles**	0.0001	**Initial principal debt amount**	0.0001	**National nationality flag**	0.0000
GDP FTEE annual deciles	0.0001	**Region**	0.0001	**Region**	0.0001	**Monthly debt amount grouped**	0.0000

Youngest DOB	0.0000	**GDP FTEE annual deciles**	0.0001	**National nationality flag**	0.0001	**# mobile phone numbers**	0.0000
Calls ever flag	0.0000	**UCPI monthly deciles**	0.0001	**Initial total debt amount**	0.0000	**GDP FTEE quarter deciles**	0.0000
Initial principal debt amount	0.0000	**Creditor grouped**	0.0000	**Calls l90d flag**	0.0000	**Product type grouped**	0.0000
GCPI annual deciles	0.0000	**Initial total debt amount**	0.0000	**CCI actual**	0.0000	**GDP MP annual deciles**	0.0000
UCPI annual deciles	0.0000	**Monthly debt amount grouped**	0.0000	**Activity rate annual**	0.0000	**GCPI monthly**	0.0000
Monthly debt amount grouped	0.0000	**Calls ever flag**	0.0000	**CCI deciles**	0.0000	**GDP FTEE annual deciles**	0.0000
Product type grouped	0.0000			**GCPI annual deciles**	0.0000	**GCPI annual deciles**	0.0000
				GDP FTEE quarter	0.0000	**UCPI annual**	0.0000
				UCPI monthly deciles	0.0000	**Initial total debt amount**	0.0000
				Calls ever flag	0.0000	**UCPI monthly**	0.0000
				Creditor grouped	0.0000	**Calls ever flag**	0.0000
						GCPI annual	0.0000

Source: Own elaboration

Table 275. Model variables and their weight for SMEs complete with payers models

Global		Pre-covid		Covid		Post-covid	
Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight
Payment l90d	0.9399	**Payment l90d**	0.9293	**Payment l90d**	0.8152	**Payment l90d**	0.8719
Payment ever	0.0308	**Payment ever**	0.0565	**Payment ever**	0.1701	**Payment ever**	0.1025
Judicial flag	0.0206	**Judicial flag**	0.0064	**Guarantor flag**	0.0068	**Calls l90d flag**	0.0157
Guarantor flag	0.0050	**Guarantor flag**	0.0044	**CCI expectations deciles**	0.0023	**Guarantor flag**	0.0058
Product type grouped	0.0007	**Monthly debt amount grouped**	0.0014	**Initial principal debt amount**	0.0020	**RPC ever flag**	0.0016
GDP MP annual	0.0003	**Calls ever flag**	0.0010	**Product type grouped**	0.0009	**Judicial flag**	0.0008

Initial principal debt amount	0.0003	**CCI expectations deciles**	0.0003	**Unemployment rate annual deciles**	0.0007	**# phone numbers**	0.0003
Creditor grouped	0.0003	**Region**	0.0001	**Initial interest debt amount**	0.0003	**# fixed line phone numbers**	0.0002
Activity rate annual	0.0003	**GCPI annual**	0.0001	**GDP FTEE annual deciles**	0.0003	**Youngest date of company creation**	0.0002
Monthly debt amount grouped	0.0002	**Creditor grouped**	0.0001	**# fixed line phone numbers**	0.0002	**Product type grouped**	0.0002
GDP MP quarter deciles	0.0002	**CCI actual**	0.0001	**GDP FTEE quarter**	0.0002	**Initial interest debt amount**	0.0002
# phone numbers	0.0002	**GDP FTEE annual**	0.0001	**UCPI monthly deciles**	0.0002	**Initial principal debt amount**	0.0002
Initial total debt amount	0.0001	**CCI actual deciles**	0.0001	**Unemployment rate annual**	0.0002	**GCPI annual**	0.0001
CCI actual	0.0001	**UCPI annual deciles**	0.0001	**GCPI annual deciles**	0.0001	**Activity rate annual**	0.0001
Calls ever flag	0.0001	**# phone numbers**	0.0000	**GDP MP quarter deciles**	0.0001	**Monthly debt amount grouped**	0.0001
GCPI annual	0.0001	**Holder or active company flag**	0.0000	**GDP MP annual deciles**	0.0001	**Calls ever flag**	0.0001
UCPI annual	0.0001	**GCPI monthly**	0.0000	**# mobile phone numbers**	0.0001	**Creditor grouped**	0.0001
CCI actual deciles	0.0001	**Initial interest debt amount**	0.0000	**CCI actual deciles**	0.0001	**UCPI annual**	0.0000
# mobile phone numbers	0.0001	**Product type grouped**	0.0000	**Oldest date of company creation**	0.0001	**GCPI monthly**	0.0000
GDP FTEE quarter deciles	0.0001			**Region**	0.0000		
GCPI annual deciles	0.0001			**Judicial flag**	0.0000		
Youngest DOB	0.0001			**CCI**	0.0000		
GDP FTEE quarter	0.0000			**Creditor grouped**	0.0000		

Source: Own elaboration

5.2.2.- Individuals Global model (complete with payers)

Variables included in the model and their importance as predictor

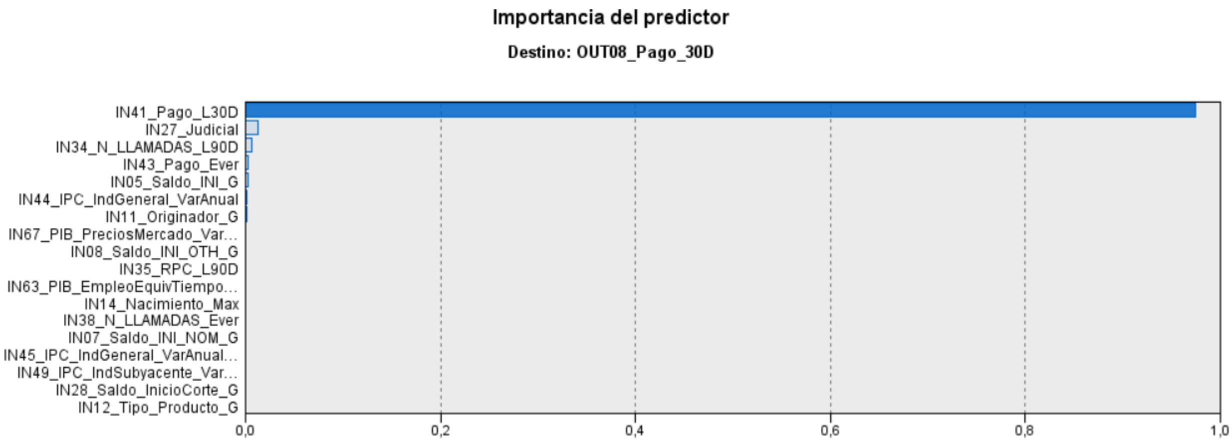
The variables included in this model are the following 18 (ordered in descendent importance) in Table 276.

Table 276. Variables included in the Individuals Global with payers model (complete)

Variable	Description
"IN41_Pago_L30D"	Payment in the last 30 days flag
"IN27_Judicial"	Judicial flag
"IN34_N_LLAMADAS_L90D"	Calls in the last 90 days flag
"IN43_Pago_Ever"	Payment ever flag
"IN05_Saldo_INI_G"	Initial total debt amount
"IN44_IPC_IndGeneral_VarAnual"	General Consumer Price Index on annual variation
"IN11_Originador_G"	Creditor grouped
"IN67_PIB_PreciosMercado_VarAnual_TIL10"	GDP Market Prices on annual variation grouped in deciles
"IN08_Saldo_INI_OTH_G"	Initial other debt amount
"IN35_RPC_L90D"	Right-party contact in the last 90 days flag
"IN63_PIB_EmpleoEquivTiempoCompleto_VarAnual_TIL10"	GDP Full-Time Equivalent Employment on annual variation grouped in deciles
"IN14_Nacimiento_Max"	Youngest date of birth or date of company creation of participants
"IN38_N_LLAMADAS_Ever"	Calls ever flag
"IN07_Saldo_INI_NOM_G"	Initial principal debt amount
"IN45_IPC_IndGeneral_VarAnual_TIL10"	General Consumer Price Index on annual variation grouped in deciles
"IN49_IPC_IndSubyacente_VarAnual_TIL10"	Underlying Consumer Price Index on annual variation grouped in deciles
"IN28_Saldo_InicioCorte_G"	Monthly debt amount grouped
"IN12_Tipo_Producto_G"	Product type grouped

Source: Own elaboration

Figure 274. Variables importance for the Individuals Global with payers model (complete)

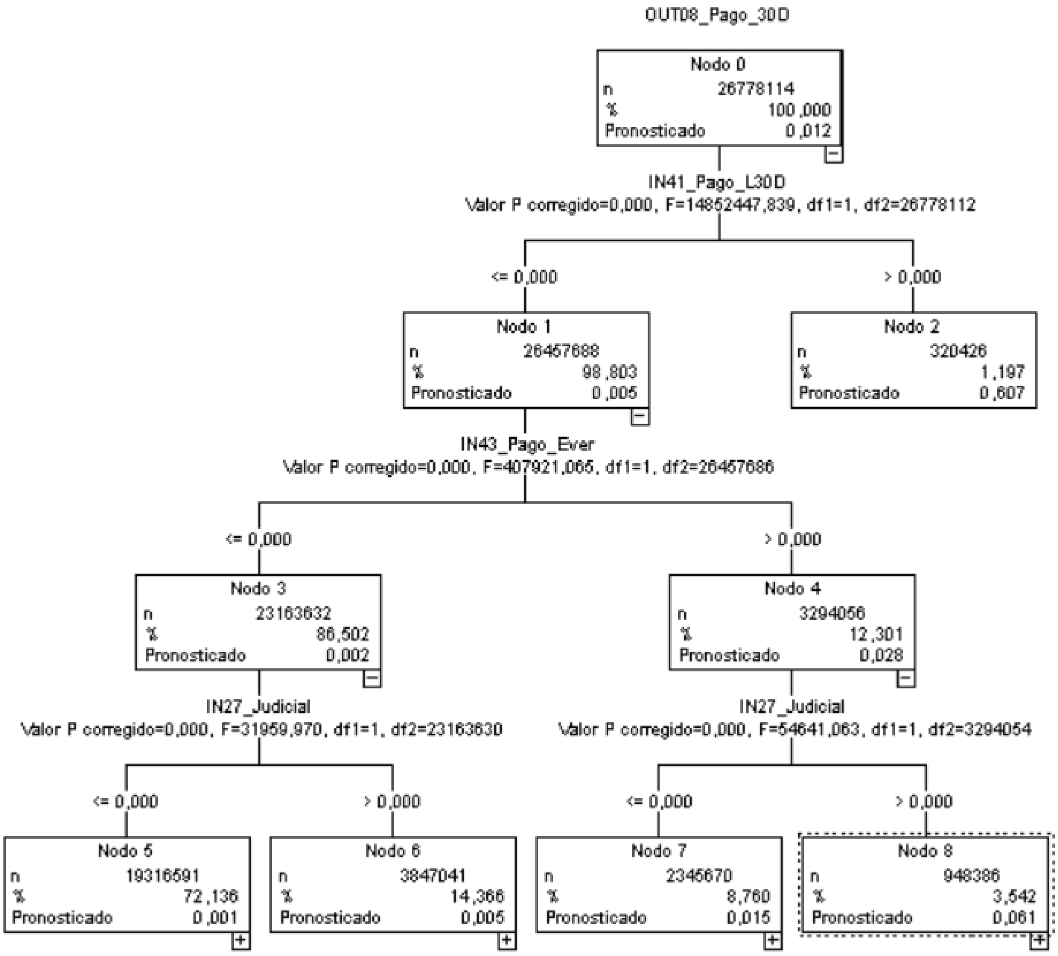


Source: Own elaboration

Model snapshot – initial layers

In the following figure, the initial 3 layers of this model are shown.

Figure 275. Initial layers of the Individuals Global with payers model (complete)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table we summarize the Gini and KS indicators for the Individuals Global model.

Table 277. Gini and K-S indicators of the Individuals Global with payers model (complete)

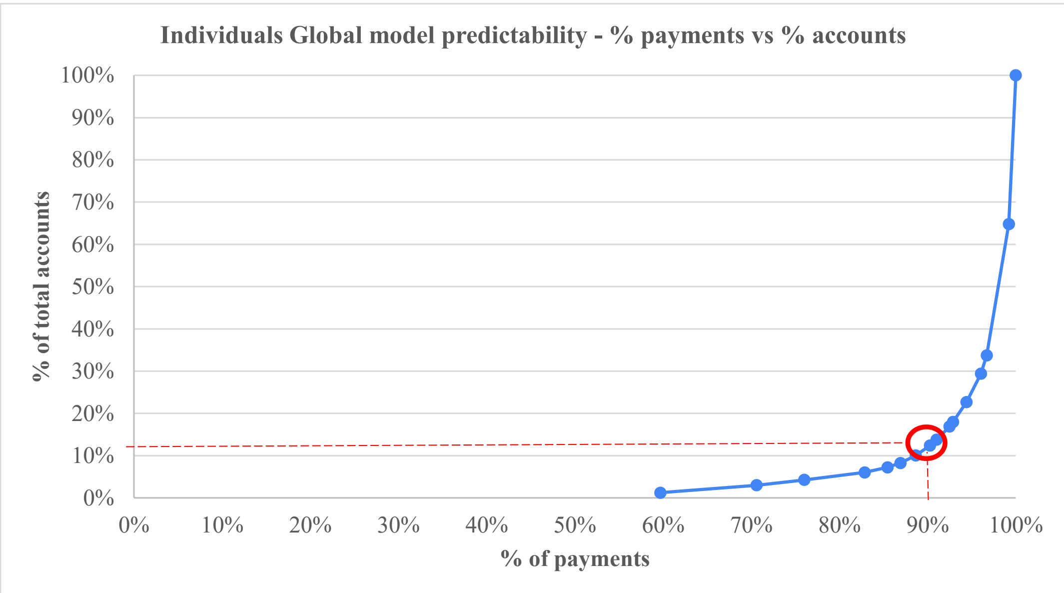
Indicators	Training sample	Testing sample
Gini	0,903	0,904
K-S	0,787	0,787

Source: Own elaboration

Model predictability: % of accounts and % of payments

Based on the model built for the Individuals Global, with a selection of 12% of the accounts, we obtain 90% of the total accounts which will pay in the following 30 days.

Figure 276. Distribution of accounts and payments on the Individuals Global with payers model (complete)



Source: Own elaboration

Table 278. Distribution of accounts and payments on the Individuals Global with payers model (complete)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	99%	65%
0,002	97%	34%
0,003	96%	29%
0,004	94%	23%
0,005	93%	18%
0,006	92%	17%
0,007	91%	14%
0,008	90%	12%
0,012	89%	10%
0,017	87%	8%
0,027	85%	7%
0,047	83%	6%
0,051	76%	4%
0,075	71%	3%
0,607	60%	1%

Source: Own elaboration

5.2.3.- Individuals Pre-Covid model (complete with payers)

Variables included in the model and their importance as predictor

The variables included in this model are the following 17 (ordered in descendent importance) in Table 279.

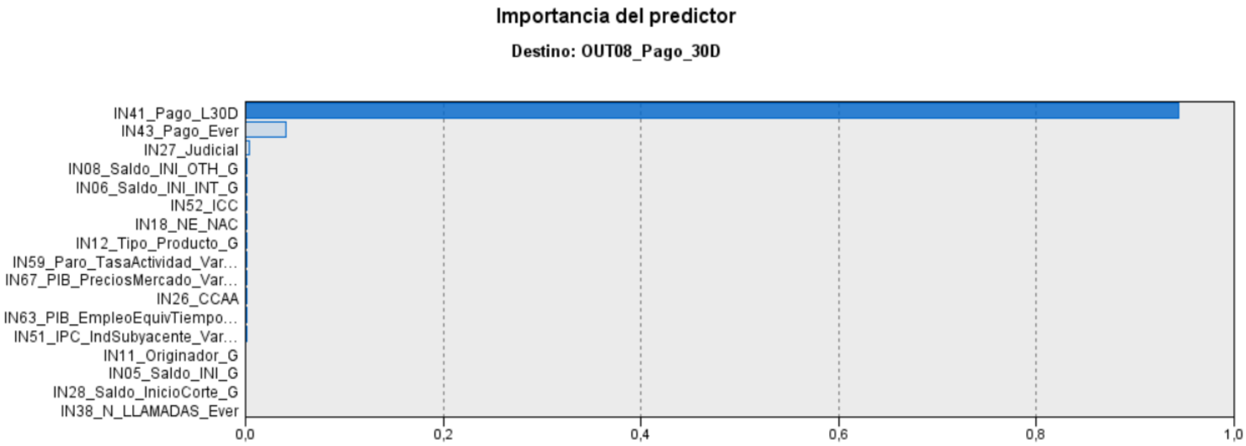
Table 279. Variables included in the Individuals Pre-Covid with payers model (complete)

Variable	Description
"IN41_Pago_L30D"	Payment in the last 30 days flag
"IN43_Pago_Ever"	Payment ever flag
"IN27_Judicial"	Judicial flag
"IN08_Saldo_INI_OTH_G"	Initial other debt amount
"IN06_Saldo_INI_INT_G"	Initial interest debt amount
"IN52_ICC"	Consumer Confidence Index
"IN18_NE_NAC"	National nationality flag
"IN12_Tipo_Producto_G"	Product type grouped
"IN59_Paro_TasaActividad_VarAnual_TIL10"	Activity Rate on annual variation grouped in deciles
"IN67_PIB_PreciosMercado_VarAnual_TIL10"	GDP Market Prices on annual variation grouped in deciles
"IN26_CCAA"	Region

"IN63_PIB_EmpleoEquivTiempoCompleto_VarAnual_TIL10"	GDP Full-Time Equivalent Employment on annual variation grouped in deciles
"IN51_IPC_IndSubyacente_VarMensual_TIL10"	Underlying Consumer Price Index on monthly variation grouped in deciles
"IN11_Originador_G"	Creditor grouped
"IN05_Saldo_INI_G"	Initial total debt amount
"IN28_Saldo_InicioCorte_G"	Monthly debt amount grouped
"IN38_N_LLAMADAS_Ever"	Calls ever flag

Source: Own elaboration

Figure 277. Variables importance for the Individuals Pre-Covid with payers model (complete)

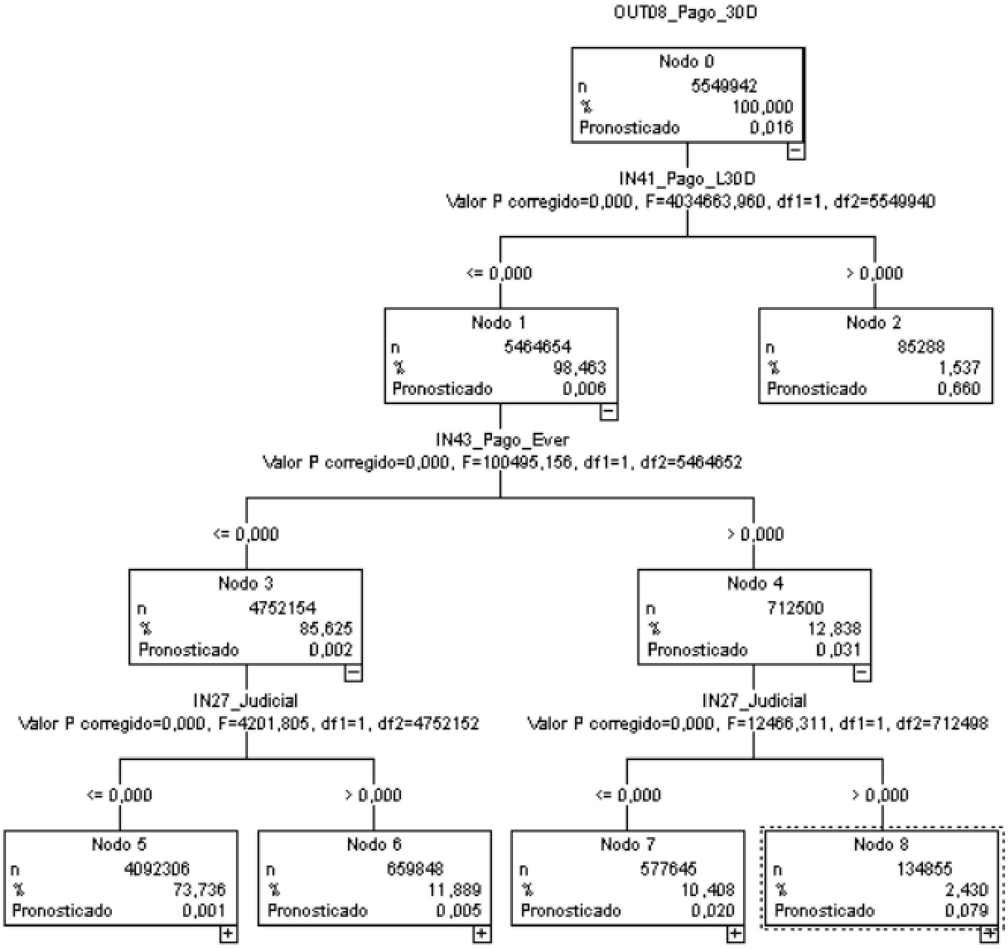


Source: Own elaboration

Model snapshot – initial layers

In the following figure, the initial 3 layers of this model are shown.

Figure 278. Initial layers of the Individuals Pre-Covid with payers model (complete)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table we summarize the Gini and KS indicators for the Individuals Pre-Covid model.

Table 280. Gini and K-S indicators of the Individuals Pre-Covid with payers model (complete)

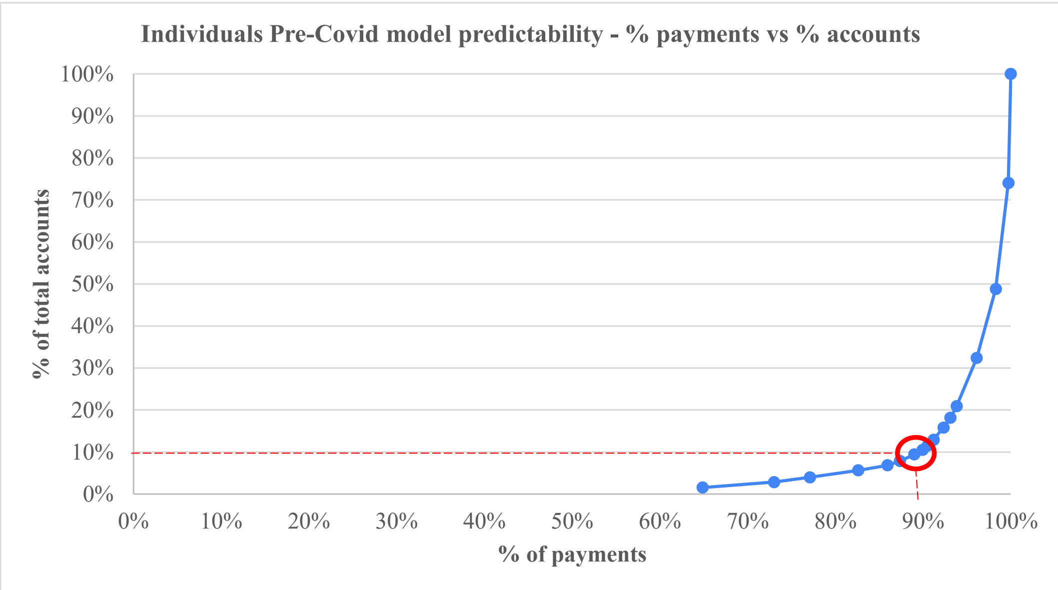
Indicators	Training sample	Testing sample
Gini	0,905	0,905
K-S	0,796	0,798

Source: Own elaboration

Model predictability: % of accounts and % of payments

Based on the model built for the Individuals Pre-Covid, with a selection of 10% of the accounts, we obtain 90% of the total accounts which will pay in the following 30 days.

Figure 279. Distribution of accounts and payments on the Individuals Pre-Covid with payers model (complete)



Source: Own elaboration

Table 281. Distribution of accounts and payments on the Individuals Pre-Covid with payers model (complete)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	100%	74%
0,002	98%	49%
0,003	96%	32%
0,004	94%	21%
0,005	93%	18%
0,006	92%	16%
0,007	91%	13%
0,009	91%	12%
0,014	90%	10%
0,017	89%	9%
0,022	87%	8%
0,044	86%	7%
0,051	83%	6%
0,057	77%	4%
0,097	73%	3%
0,660	65%	2%

Source: Own elaboration

5.2.4.- Individuals Covid model (complete with payers)

Variables included in the model and their importance as predictor

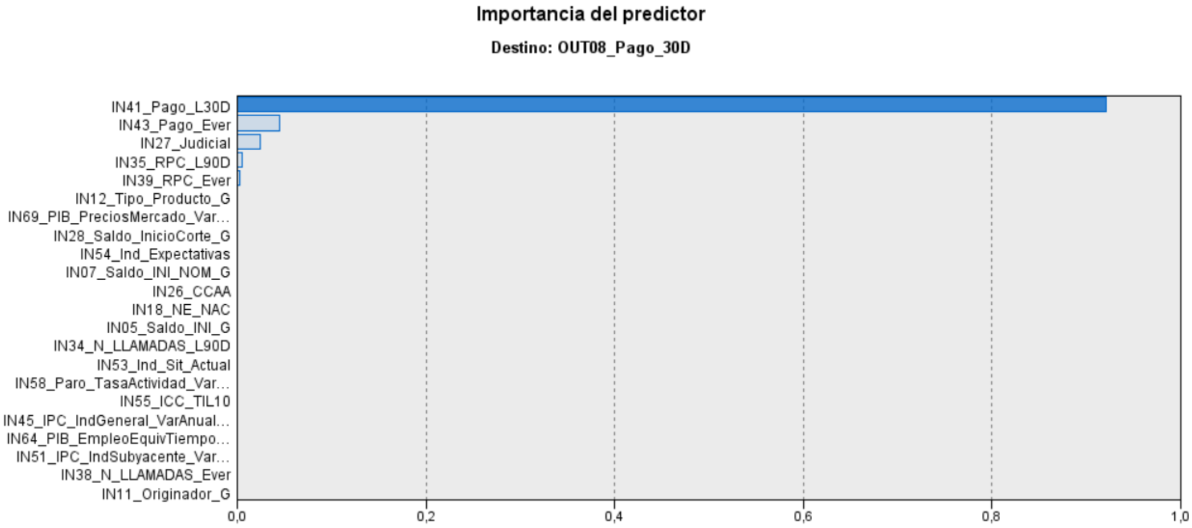
The variables included in this model are the following 22 (ordered in descendent importance) in Table 282.

Table 282. Variables included in the Individuals Covid with payers model (complete)

Variable	Description
"IN41_Pago_L30D"	Payment in the last 30 days flag
"IN43_Pago_Ever"	Payment ever flag
"IN27_Judicial"	Judicial flag
"IN35_RPC_L90D"	Right-party contact in the last 90 days flag
"IN39_RPC_Ever"	Right-party contact ever flag
"IN12_Tipo_Producto_G"	Product type grouped
"IN69_PIB_PreciosMercado_VarTrim_TIL10"	GDP Market Prices on quarterly variation grouped in deciles
"IN28_Saldo_InicioCorte_G"	Monthly debt amount grouped
"IN54_Ind_Expectativas"	Consumer Confidence Index expectations
"IN07_Saldo_INI_NOM_G"	Initial principal debt amount
"IN26_CCAA"	Region
"IN18_NE_NAC"	National nationality flag
"IN05_Saldo_INI_G"	Initial total debt amount
"IN34_N_LLAMADAS_L90D"	Calls in the last 90 days flag
"IN53_Ind_Sit_Actual"	Consumer Confidence Index actual situation
"IN58_Paro_TasaActividad_VarAnual"	Activity Rate on annual variation
"IN55_ICC_TIL10"	Consumer Confidence Index grouped in deciles
"IN45_IPC_IndGeneral_VarAnual_TIL10"	General Consumer Price Index on annual variation grouped in deciles
"IN64_PIB_EmpleoEquivTiempoCompleto_VarTrim"	GDP Full-Time Equivalent Employment on quarterly variation
"IN51_IPC_IndSubyacente_VarMensual_TIL10"	Underlying Consumer Price Index on monthly variation grouped in deciles
"IN38_N_LLAMADAS_Ever"	Calls ever flag
"IN11_Originador_G"	Creditor grouped

Source: Own elaboration

Figure 280. Variables importance for the Individuals Covid with payers model (complete)

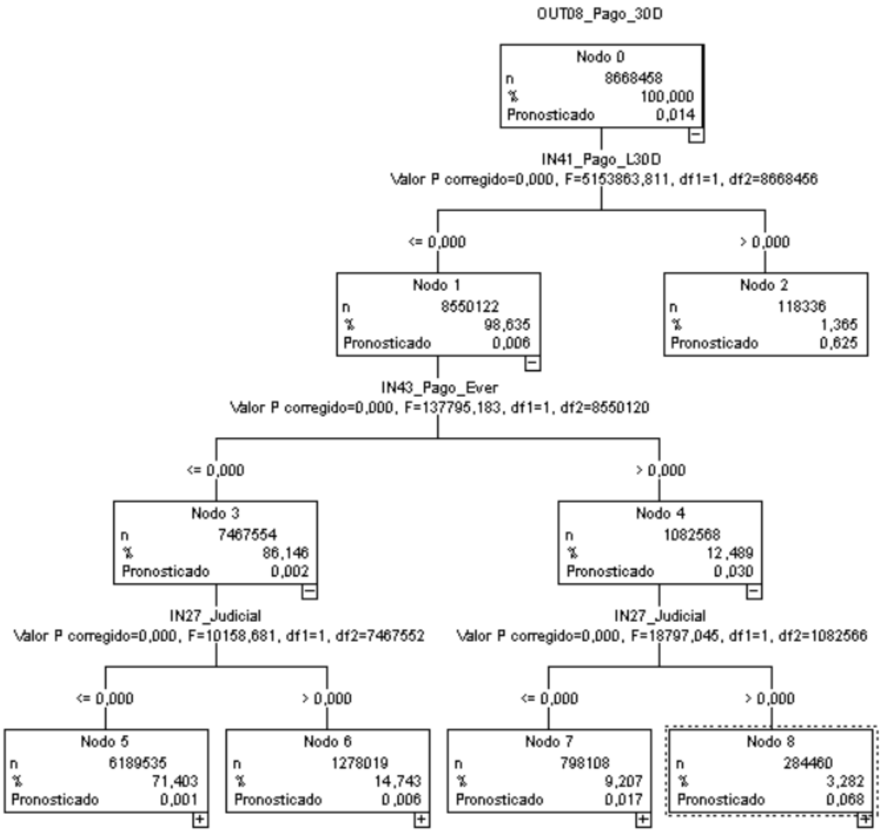


Source: Own elaboration

Model snapshot – initial layers

In the following figure, the initial 3 layers of this model are shown.

Figure 281. Initial layers of the Individuals Covid with payers model (complete)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table we summarize the Gini and KS indicators for the Individuals Covid model.

Table 283. Gini and K-S indicators of the Individuals Covid with payers model (complete)

Indicators	Training sample	Testing sample
Gini	0,897	0,896
K-S	0,780	0,778

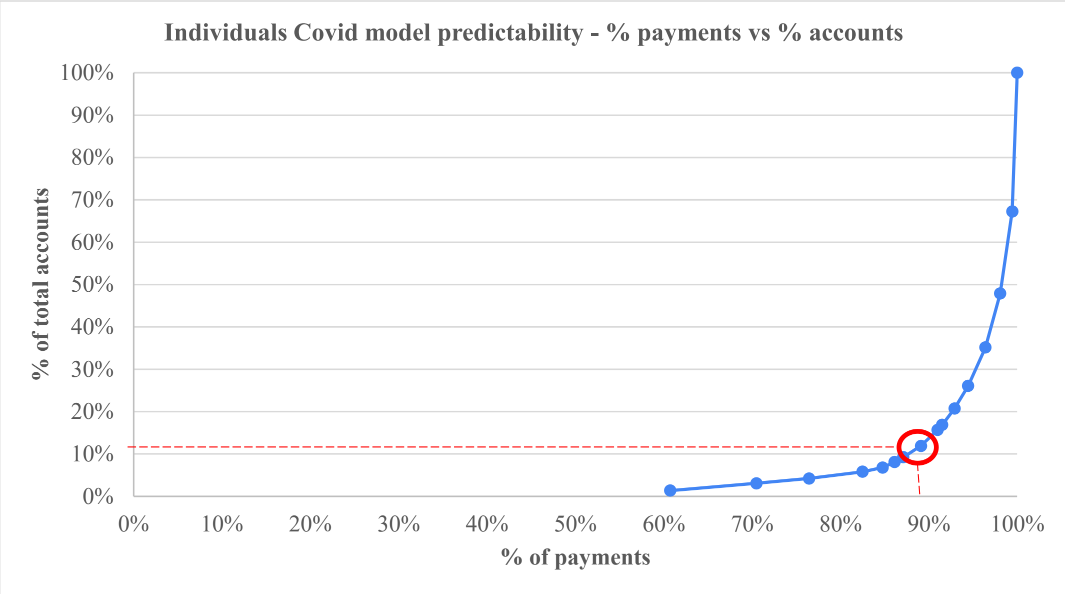
Source: Own elaboration

Model predictability: % of accounts and % of payments

Based on the model built for the Individuals Covid, with a selection of 12% of the accounts, we obtain 89% of the total accounts which will pay in the following 30 days.

Table 284 and Figure 282. Distribution of accounts and payments on the Individuals Covid with payers model (complete)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	99%	67%
0,002	98%	48%
0,003	96%	35%
0,004	94%	26%
0,005	93%	21%
0,006	92%	17%
0,007	91%	16%
0,011	89%	12%
0,012	87%	9%
0,014	86%	8%
0,032	85%	7%
0,054	82%	6%
0,073	76%	4%
0,081	70%	3%
0,625	61%	1%



Source: Own elaboration

5.2.5.- Individuals Post-Covid model (complete with payers)

Variables included in the model and their importance as predictor

The variables included in this model are the following 23 (ordered in descendent importance) in Table 285.

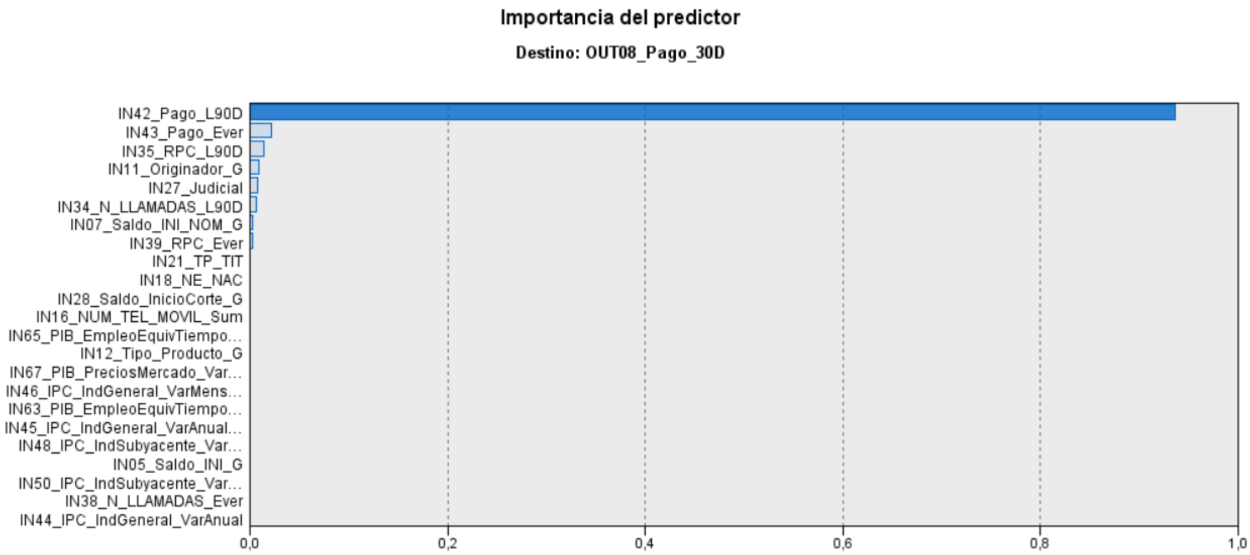
Table 285. Variables included in the Individuals Post-Covid with payers model (complete)

Variable	Description
"IN42_Pago_L90D"	Payment in the last 90 days flag
"IN43_Pago_Ever"	Payment ever flag
"IN35_RPC_L90D"	Right-party contact in the last 90 days flag
"IN11_Originador_G"	Creditor grouped
"IN27_Judicial"	Judicial flag
"IN34_N_LLAMADAS_L90D"	Calls in the last 90 days flag
"IN07_Saldo_INI_NOM_G"	Initial principal debt amount
"IN39_RPC_Ever"	Right-party contact ever flag
"IN21_TP_TIT"	Holder or active company flag
"IN18_NE_NAC"	National nationality flag
"IN28_Saldo_InicioCorte_G"	Monthly debt amount grouped
"IN16_NUM_TEL_MOVIL_Sum"	Number of mobile phone numbers
"IN65_PIB_EmpleoEquivTiempoCompleto_VarTrim_TIL10"	GDP Full-Time Equivalent Employment on quarterly variation grouped in deciles

“IN12_Tipo_Producto_G”	Product type grouped
“IN67_PIB_PreciosMercado_VarAnual_TIL10”	GDP Market Prices on annual variation grouped in deciles
“IN46_IPC_IndGeneral_VarMensual”	General Consumer Price Index on monthly variation
“IN63_PIB_EmpleoEquivTiempoCompleto_VarAnual_TIL10”	GDP Full-Time Equivalent Employment on annual variation grouped in deciles
“IN45_IPC_IndGeneral_VarAnual_TIL10”	General Consumer Price Index on annual variation grouped in deciles
“IN48_IPC_IndSubyacente_VarAnual”	Underlying Consumer Price Index on annual variation
“IN05_Saldo_INI_G”	Initial total debt amount
“IN50_IPC_IndSubyacente_VarMensual”	Underlying Consumer Price Index on monthly variation
“IN38_N_LLAMADAS_Ever”	Calls ever flag
“IN44_IPC_IndGeneral_VarAnual”	General Consumer Price Index on annual variation

Source: Own elaboration

Figure 283. Variables importance for the Individuals Post-Covid with payers model (complete)

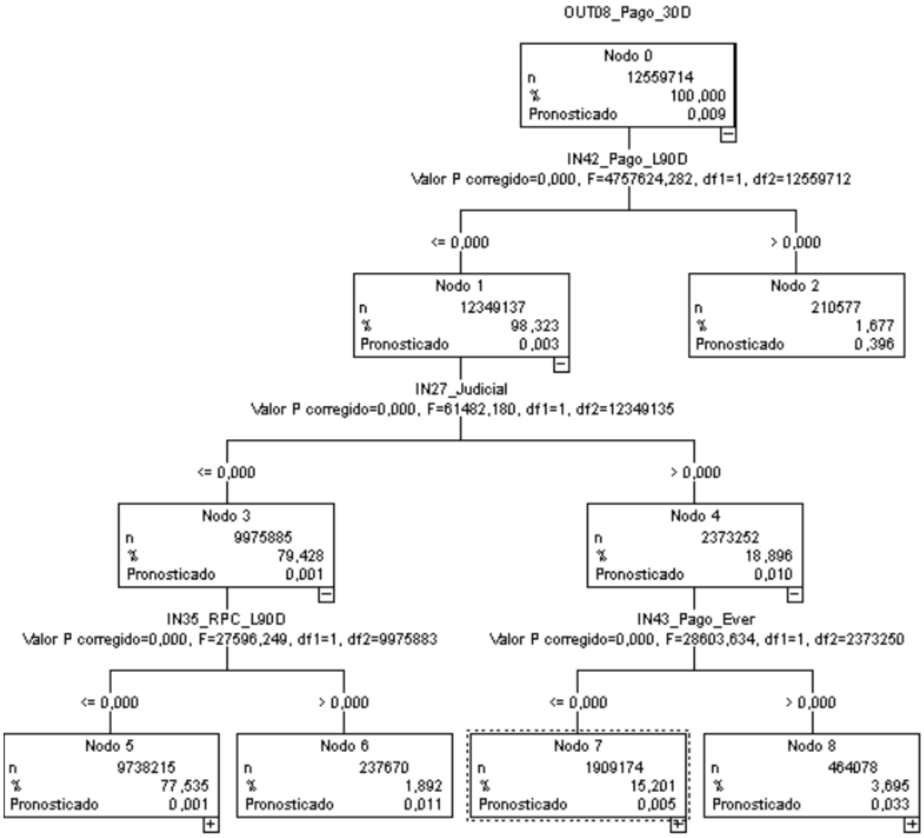


Source: Own elaboration

Model snapshot – initial layers

In the following figure, the initial 3 layers of this model are shown.

Figure 284. Initial layers of the Individuals Post-Covid with payers model (complete)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table we summarize the Gini and KS indicators for the Individuals Post-Covid model.

Table 286. Gini and K-S indicators of the Individuals Post-Covid with payers model (complete)

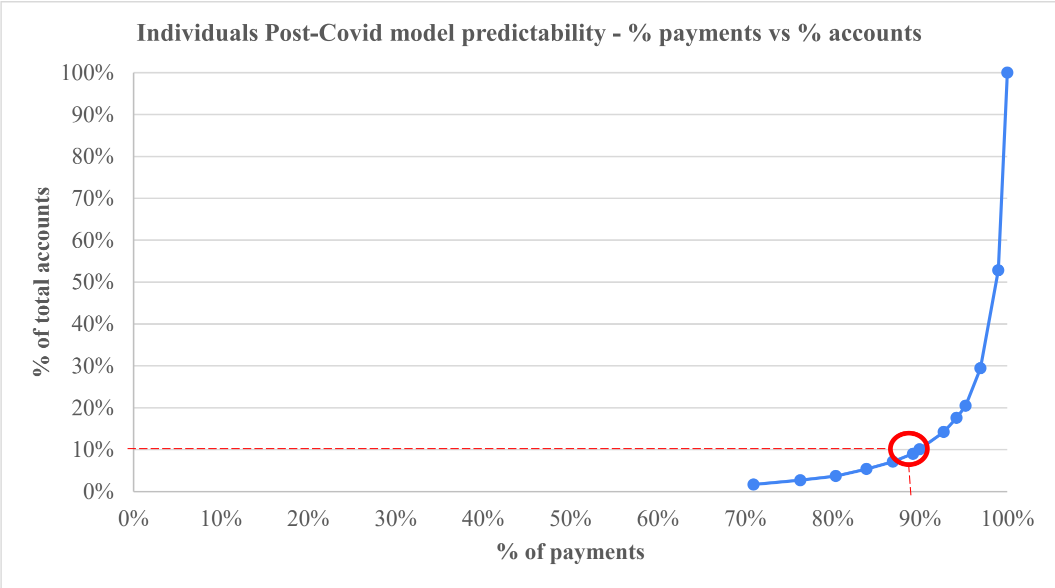
Indicators	Training sample	Testing sample
Gini	0,917	0,917
K-S	0,802	0,802

Source: Own elaboration

Model predictability: % of accounts and % of payments

Based on the model built for the Individuals Post-Covid, with a selection of 10% of the accounts, we obtain 90% of the total accounts which will pay in the following 30 days.

Figure 285. Distribution of accounts and payments on the Individuals Pre-Covid with payers model (complete)



Source: Own elaboration

Table 287. Distribution of accounts and payments on the Individuals Pre-Covid with payers model (complete)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	99%	53%
0,002	97%	29%
0,003	95%	20%
0,004	94%	18%
0,006	93%	14%
0,007	90%	10%
0,011	89%	9%
0,016	87%	7%
0,020	84%	5%
0,038	80%	4%
0,050	76%	3%
0,396	71%	2%

Source: Own elaboration

5.2.6.- SMEs Global model (complete with payers)

Variables included in the model and their importance as predictor

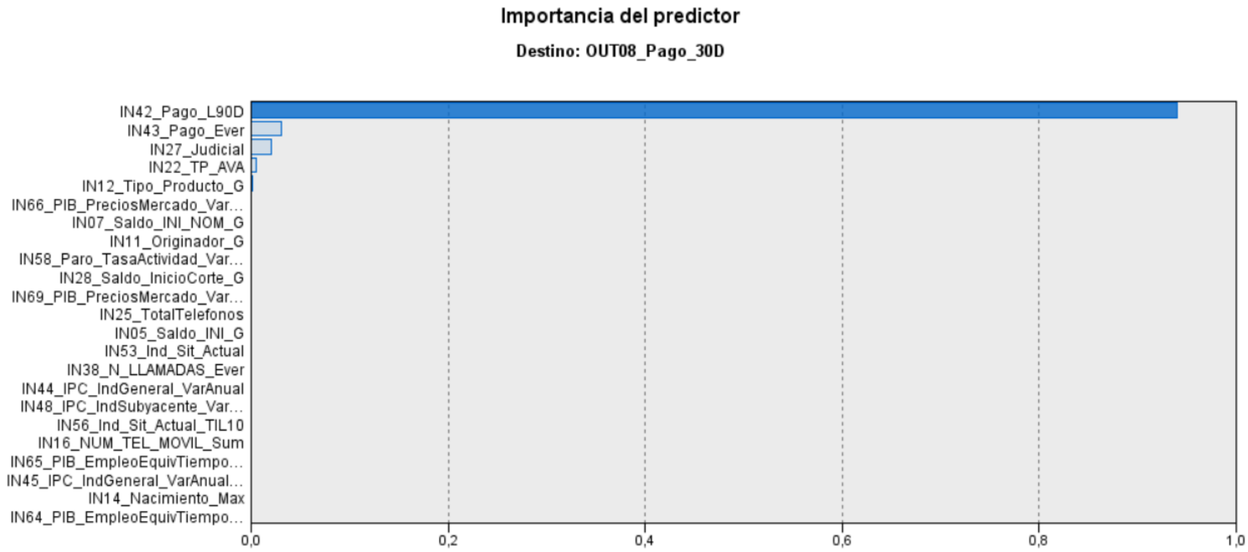
The variables included in this model are the following 23 (ordered in descendent importance) in Table 288.

Table 288. Variables included in the SMEs Global with payers model (complete)

Variable	Description
"IN42_Pago_L90D"	Payment in the last 90 days flag
"IN43_Pago_Ever"	Payment ever flag
"IN27_Judicial"	Judicial flag
"IN22_TP_AVA"	Guarantor flag
"IN12_Tipo_Producto_G"	Product type grouped
"IN66_PIB_PreciosMercado_VarAnual"	GDP Market Prices on annual variation
"IN07_Saldo_INI_NOM_G"	Initial principal debt amount
"IN11_Originador_G"	Creditor grouped
"IN58_Paro_TasaActividad_VarAnual"	Activity Rate on annual variation
"IN28_Saldo_InicioCorte_G"	Monthly debt amount grouped
"IN69_PIB_PreciosMercado_VarTrim_TIL10"	GDP Market Prices on quarterly variation grouped in deciles
"IN25_TotalTelefonos"	Number of phone numbers
"IN05_Saldo_INI_G"	Initial total debt amount
"IN53_Ind_Sit_Actual"	Consumer Confidence Index actual situation
"IN38_N_LLAMADAS_Ever"	Calls ever flag
"IN44_IPC_IndGeneral_VarAnual"	General Consumer Price Index on annual variation
"IN48_IPC_IndSubyacente_VarAnual"	Underlying Consumer Price Index on annual variation
"IN56_Ind_Sit_Actual_TIL10"	Consumer Confidence Index actual situation grouped in deciles
"IN16_NUM_TEL_MOVIL_Sum"	Number of mobile phone numbers
"IN65_PIB_EmpleoEquivTiempoCompleto_VarTrim_TIL10"	GDP Full-Time Equivalent Employment on quarterly variation grouped in deciles
"IN45_IPC_IndGeneral_VarAnual_TIL10"	General Consumer Price Index on annual variation grouped in deciles
"IN14_Nacimiento_Max"	Youngest date of birth or date of company creation of participants
"IN64_PIB_EmpleoEquivTiempoCompleto_VarTrim"	GDP Full-Time Equivalent Employment on quarterly variation

Source: Own elaboration

Figure 286. Variables importance for the SMEs Global with payers model (complete)

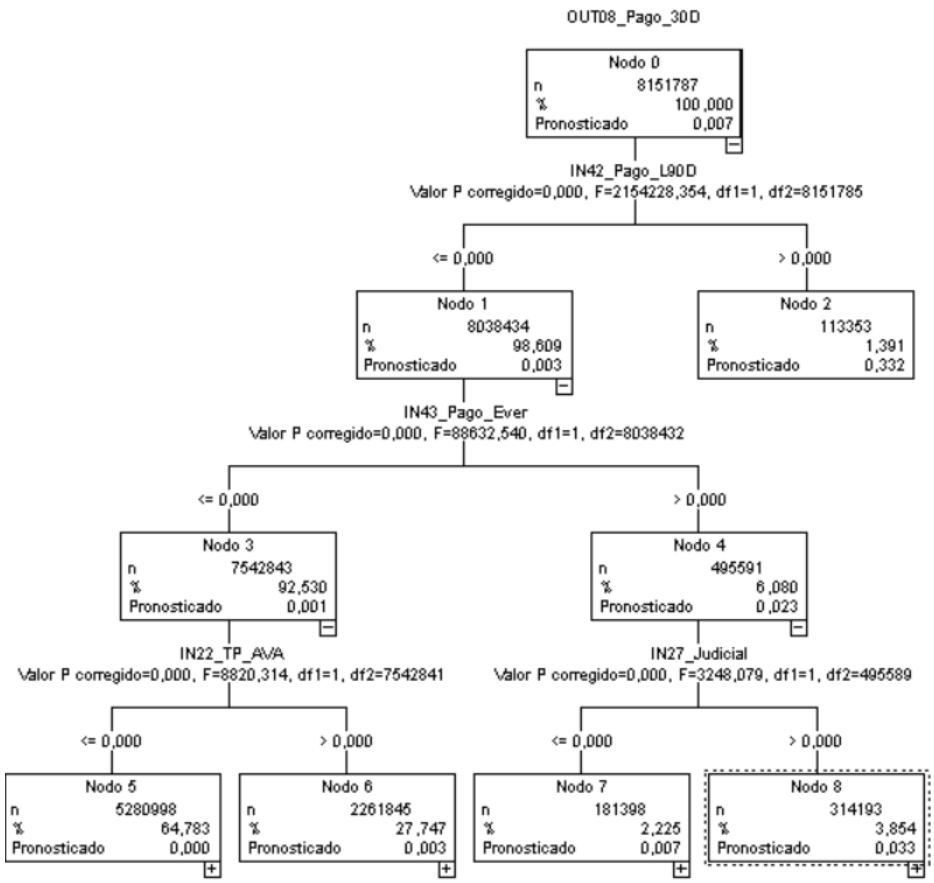


Source: Own elaboration

Model snapshot – initial layers

In the following figure, the initial 3 layers of this model are shown.

Figure 287. Initial layers of the SMEs Global with payers model (complete)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table we summarize the Gini and KS indicators for the SMEs Global model.

Table 289. Gini and K-S indicators of the SMEs Global with payers model (complete)

Indicators	Training sample	Testing sample
Gini	0,905	0,900
K-S	0,777	0,774

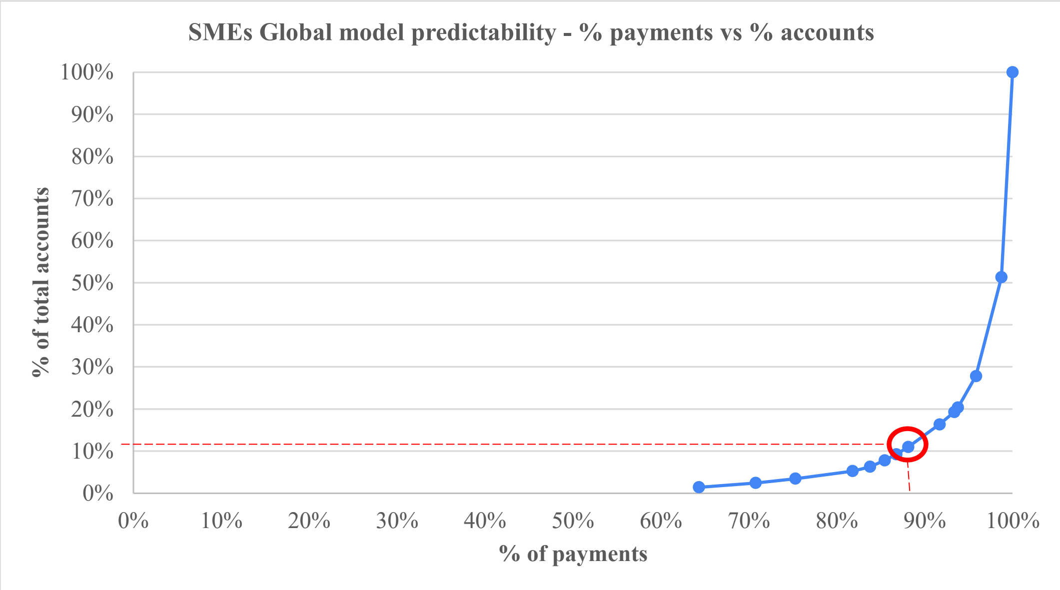
Source: Own elaboration

Model predictability: % of accounts and % of payments

Based on the model built for the SMEs Global, with a selection of 11% of the accounts, we obtain 88% of the total accounts which will pay in the following 30 days.

Table 290 and Figure 288. Distribution of accounts and payments on the SMEs Global with payers model (complete)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	99%	51%
0,002	96%	28%
0,003	94%	20%
0,004	93%	19%
0,005	92%	16%
0,006	88%	11%
0,007	87%	9%
0,008	85%	8%
0,014	84%	6%
0,026	82%	5%
0,032	75%	3%
0,045	71%	2%
0,332	64%	1%



Source: Own elaboration

5.2.7.- SMEs Pre-Covid model (complete with payers)

Variables included in the model and their importance as predictor

The variables included in this model are the following 19 (ordered in descendent importance) in Table 291.

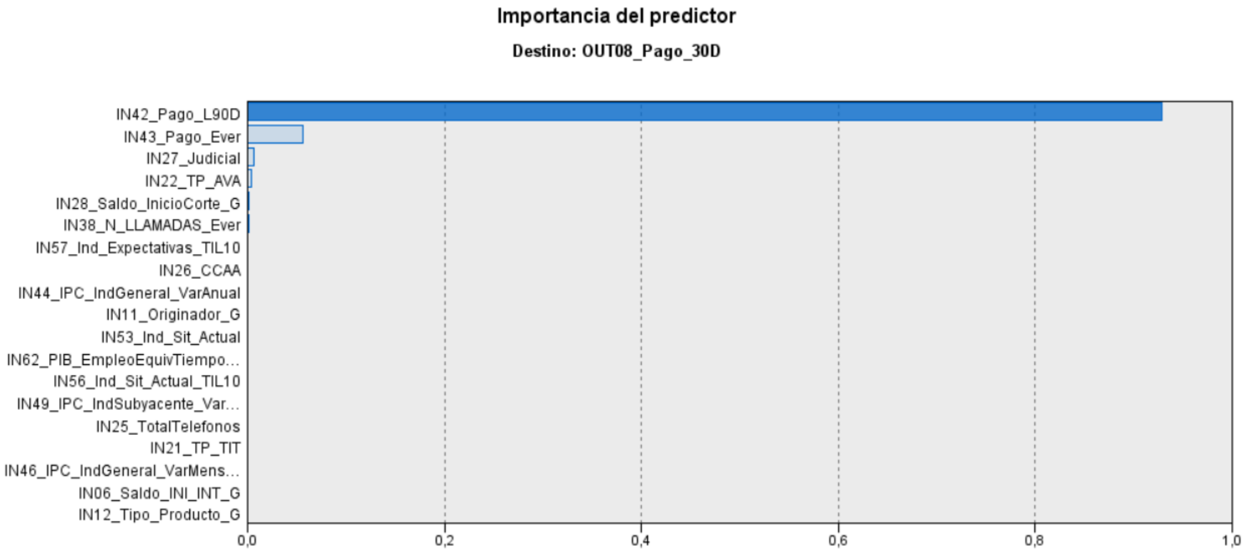
Table 291. Variables included in the SMEs Pre-Covid with payers model (complete)

Variable	Description
"IN42_Pago_L90D"	Payment in the last 90 days flag
"IN43_Pago_Ever"	Payment ever flag
"IN27_Judicial"	Judicial flag
"IN22_TP_AVA"	Guarantor flag
"IN28_Saldo_InicioCorte_G"	Monthly debt amount grouped
"IN38_N_LLAMADAS_Ever"	Calls ever flag
"IN57_Ind_Expectativas_TIL10"	Consumer Confidence Index expectations grouped in deciles
"IN26_CCAA"	Region
"IN44_IPC_IndGeneral_VarAnual"	General Consumer Price Index on annual variation
"IN11_Originador_G"	Creditor grouped
"IN53_Ind_Sit_Actual"	Consumer Confidence Index actual situation
"IN62_PIB_EmpleoEquivTiempoCompleto_VarAnual"	GDP Full-Time Equivalent Employment on annual variation
"IN56_Ind_Sit_Actual_TIL10"	Consumer Confidence Index actual situation grouped in deciles

"IN49_IPC_IndSubyacente_VarAnual_TIL10"	Underlying Consumer Price Index on annual variation grouped in deciles
"IN25_TotalTelefonos"	Number of phone numbers
"IN21_TP_TIT"	Holder or active company flag
"IN46_IPC_IndGeneral_VarMensual"	General Consumer Price Index on monthly variation
"IN06_Saldo_INI_INT_G"	Initial interest debt amount
"IN12_Tipo_Producto_G"	Product type grouped

Source: Own elaboration

Figure 289. Variables importance for the SMEs Pre-Covid with payers model (complete)

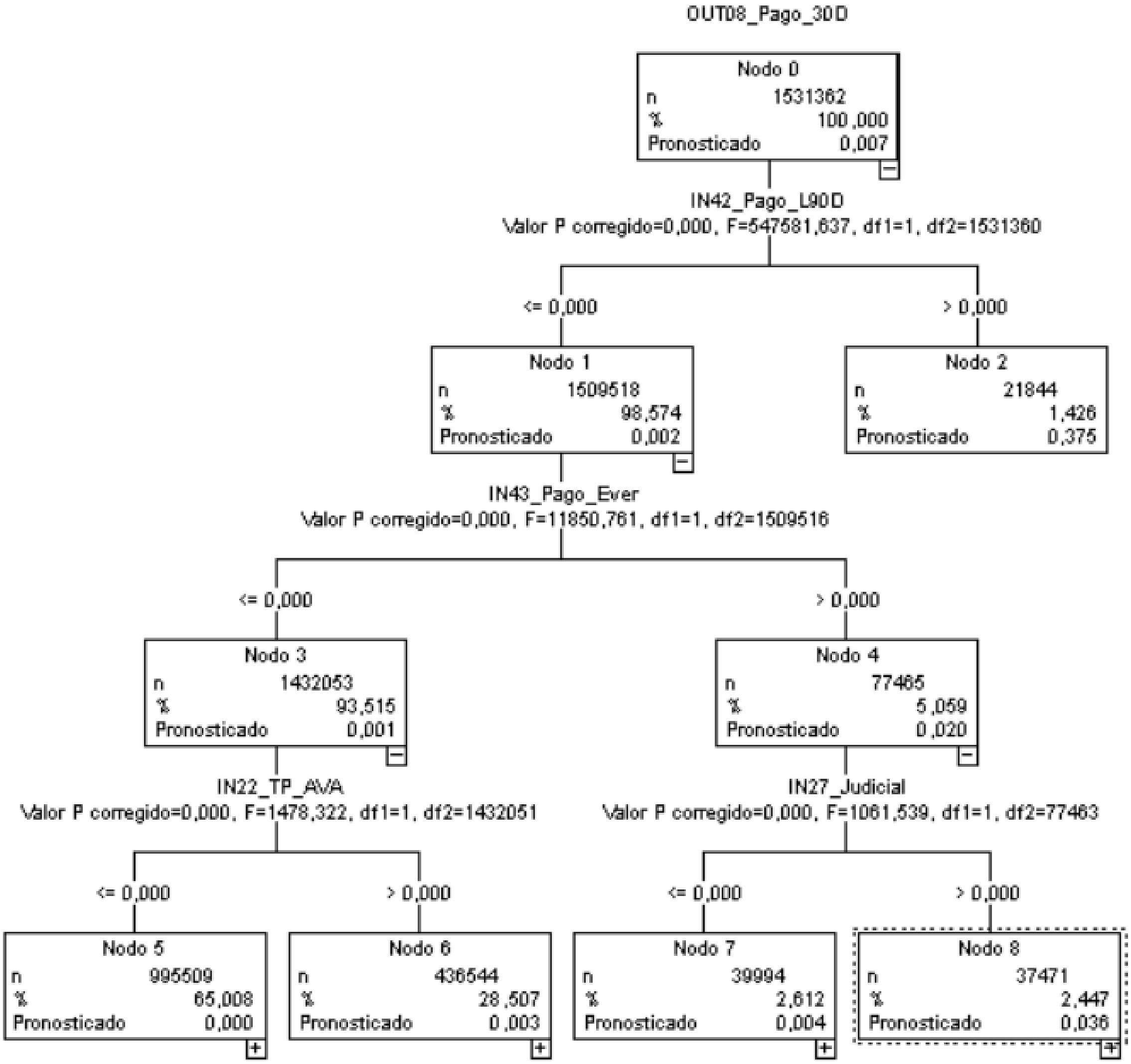


Source: Own elaboration

Model snapshot – initial layers

In the following figure, the initial 3 layers of this model are shown.

Figure 290. Initial layers of the SMEs Pre-Covid with payers model (complete)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table we summarize the Gini and KS indicators for the SMEs Pre-Covid model.

Table 291. Gini and K-S indicators of the SMEs Pre-Covid with payers model (complete)

Indicators	Training sample	Testing sample
Gini	0,920	0,908
K-S	0,798	0,792

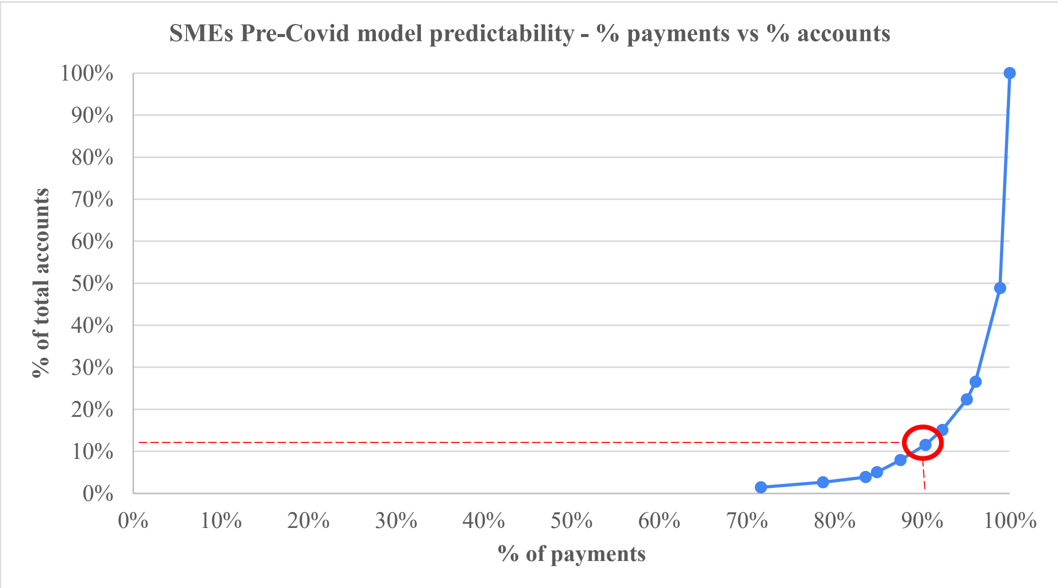
Source: Own elaboration

Model predictability: % of accounts and % of payments

Based on the model built for the SMEs Pre-Covid, with a selection of 12% of the accounts, we obtain 90% of the total accounts which will pay in the following 30 days.

Table 292 and Figure 291. Distribution of accounts and payments on the SMEs Pre-Covid with payers model (complete)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	99%	49%
0,002	96%	27%
0,003	95%	22%
0,004	92%	15%
0,006	90%	12%
0,007	88%	8%
0,009	85%	5%
0,029	84%	4%
0,044	79%	3%
0,375	72%	1%



Source: Own elaboration

5.2.8.- SMEs Covid model (complete with payers)

Variables included in the model and their importance as predictor

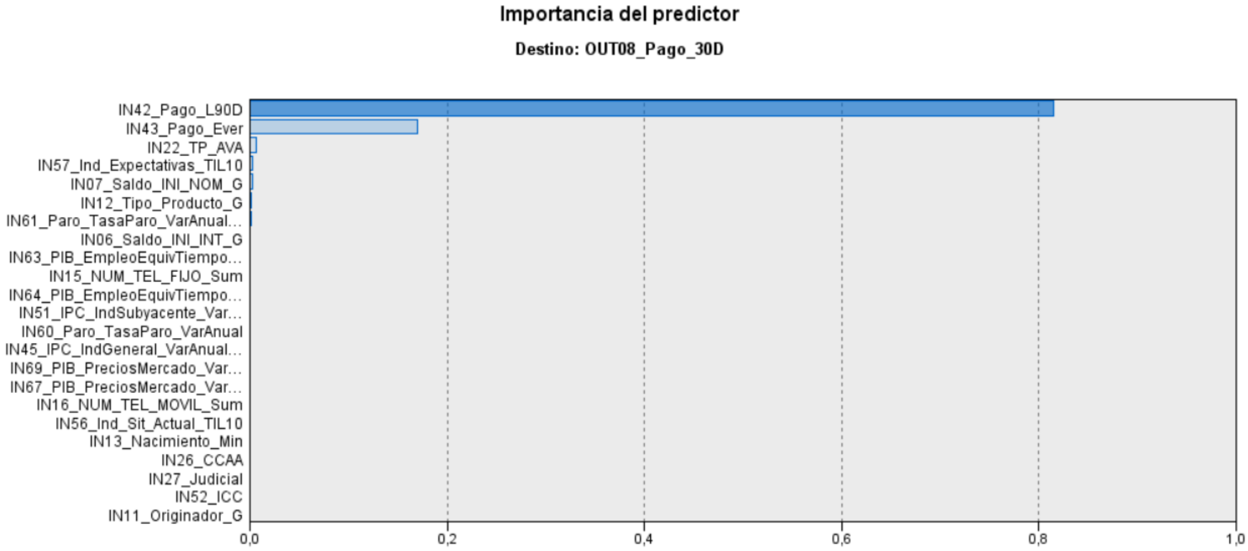
The variables included in this model are the following 23 (ordered in descendent importance) in Table 293.

Table 293. Variables included in the SMEs Covid with payers model (complete)

Variable	Description
"IN42_Pago_L90D"	Payment in the last 90 days flag
"IN43_Pago_Ever"	Payment ever flag
"IN22_TP_AVA"	Guarantor flag
"IN57_Ind_Expectativas_TIL10"	Consumer Confidence Index expectations grouped in deciles
"IN07_Saldo_INI_NOM_G"	Initial principal debt amount
"IN12_Tipo_Producto_G"	Product type grouped
"IN61_Paro_TasaParo_VarAnual_TIL10"	Unemployment Rate on annual variation grouped in deciles
"IN06_Saldo_INI_INT_G"	Initial interest debt amount
"IN63_PIB_EmpleoEquivTiempoCompleto_VarAnual_TIL10"	GDP Full-Time Equivalent Employment on annual variation grouped in deciles
"IN15_NUM_TEL_FIJO_Sum"	Number of fixed line phone numbers
"IN64_PIB_EmpleoEquivTiempoCompleto_VarTrim"	GDP Full-Time Equivalent Employment on quarterly variation
"IN51_IPC_IndSubyacente_VarMensual_TIL10"	Underlying Consumer Price Index on monthly variation grouped in deciles
"IN60_Paro_TasaParo_VarAnual"	Unemployment Rate on annual variation
"IN45_IPC_IndGeneral_VarAnual_TIL10"	General Consumer Price Index on annual variation grouped in deciles
"IN69_PIB_PreciosMercado_VarTrim_TIL10"	GDP Market Prices on quarterly variation grouped in deciles
"IN67_PIB_PreciosMercado_VarAnual_TIL10"	GDP Market Prices on annual variation grouped in deciles
"IN16_NUM_TEL_MOVIL_Sum"	Number of mobile phone numbers
"IN56_Ind_Sit_Actual_TIL10"	Consumer Confidence Index actual situation grouped in deciles
"IN13_Nacimiento_Min"	Oldest date of birth or date of company creation of participants
"IN26_CCAA"	Region
"IN27_Judicial"	Judicial flag
"IN52_ICC"	Consumer Confidence Index
"IN11_Originador_G"	Creditor grouped

Source: Own elaboration

Figure 292. Variables importance for the SMEs Covid with payers model (complete)

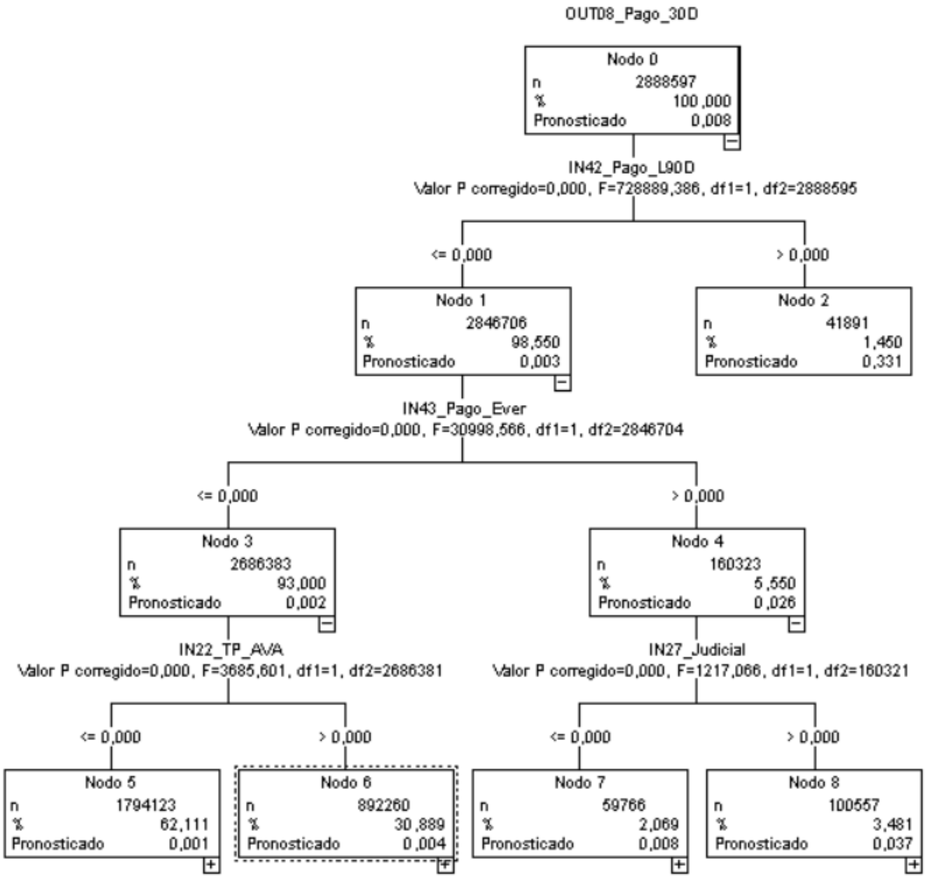


Source: Own elaboration

Model snapshot – initial layers

In the following figure, the initial 3 layers of this model are shown.

Figure 293. Initial layers of the SMEs Covid with payers model (complete)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table we summarize the Gini and KS indicators for the SMEs Covid model.

Table 294. Gini and K-S indicators of the SMEs Covid with payers model (complete)

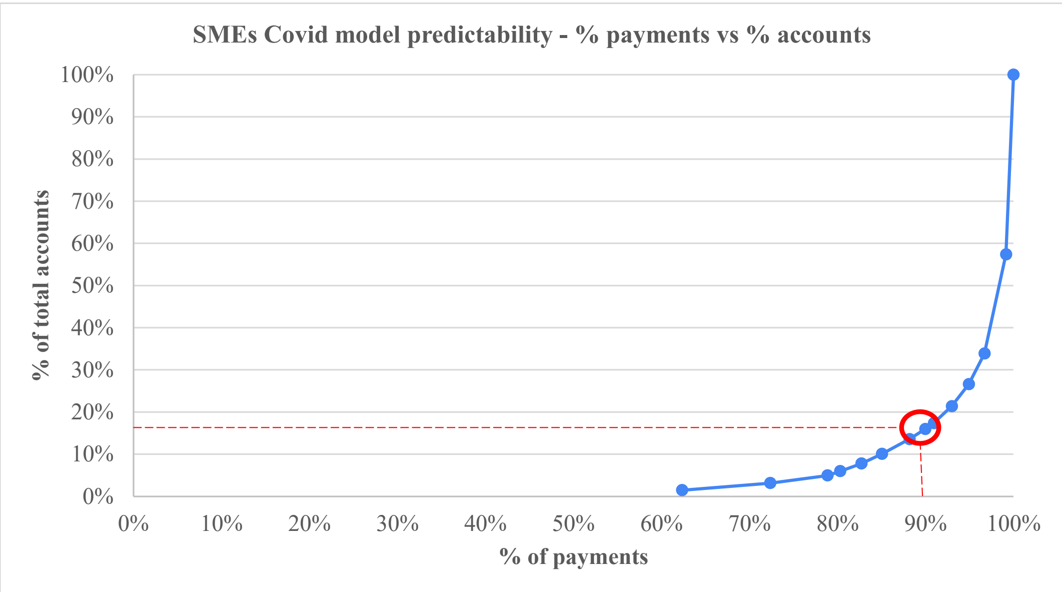
Indicators	Training sample	Testing sample
Gini	0,895	0,891
K-S	0,750	0,752

Source: Own elaboration

Model predictability: % of accounts and % of payments

Based on the model built for the SMEs Covid, with a selection of 16% of the accounts, we obtain 90% of the total accounts which will pay in the following 30 days.

Figure 294. Distribution of accounts and payments on the SMEs Covid with payers model (complete)



Source: Own elaboration

Table 295. Distribution of accounts and payments on the SMEs Covid with payers model (complete)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	99%	57%
0,002	97%	34%
0,003	95%	27%
0,004	93%	21%
0,005	91%	17%
0,006	90%	16%
0,007	88%	14%
0,008	85%	10%
0,010	83%	8%
0,011	80%	6%
0,028	79%	5%
0,046	72%	3%
0,331	62%	1%

Source: Own elaboration

5.2.9.- SMEs Post-Covid model (complete with payers)

Variables included in the model and their importance as predictor

The variables included in this model are the following 19 (ordered in descendent importance) in Table 296.

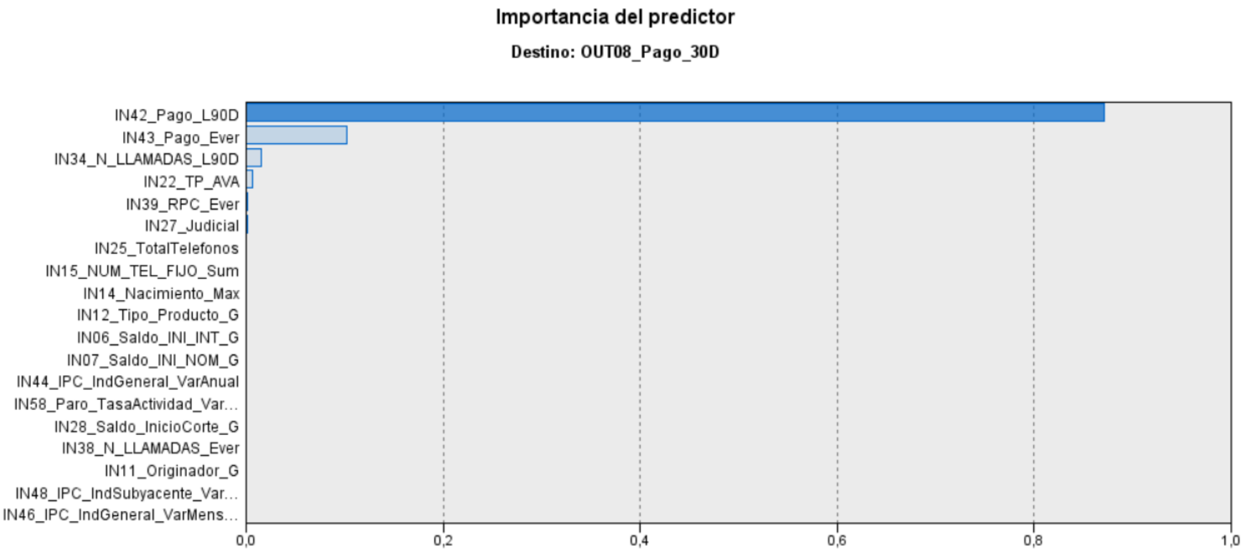
Table 296. Variables included in the SMEs Post-Covid with payers model (complete)

Variable	Description
"IN42_Pago_L90D"	Payment in the last 90 days flag
"IN43_Pago_Ever"	Payment ever flag
"IN34_N_LLAMADAS_L90D"	Calls in the last 90 days flag
"IN22_TP_AVA"	Guarantor flag
"IN39_RPC_Ever"	Right-party contact ever flag
"IN27_Judicial"	Judicial flag
"IN25_TotalTelefonos"	Number of phone numbers
"IN15_NUM_TEL_FIJO_Sum"	Number of fixed line phone numbers
"IN14_Nacimiento_Max"	Youngest date of birth or date of company creation of participants
"IN12_Tipo_Producto_G"	Product type grouped
"IN06_Saldo_INI_INT_G"	Initial interest debt amount
"IN07_Saldo_INI_NOM_G"	Initial principal debt amount
"IN44_IPC_IndGeneral_VarAnual"	General Consumer Price Index on annual variation
"IN58_Paro_TasaActividad_VarAnual"	Activity Rate on annual variation

"IN28_Saldo_InicioCorte_G"	Monthly debt amount grouped
"IN38_N_LLAMADAS_Ever"	Calls ever flag
"IN11_Originador_G"	Creditor grouped
"IN48_IPC_IndSubyacente_VarAnual"	Underlying Consumer Price Index on annual variation
"IN46_IPC_IndGeneral_VarMensual"	General Consumer Price Index on monthly variation

Source: Own elaboration

Figure 295. Variables importance for the SMEs Post-Covid with payers model (complete)

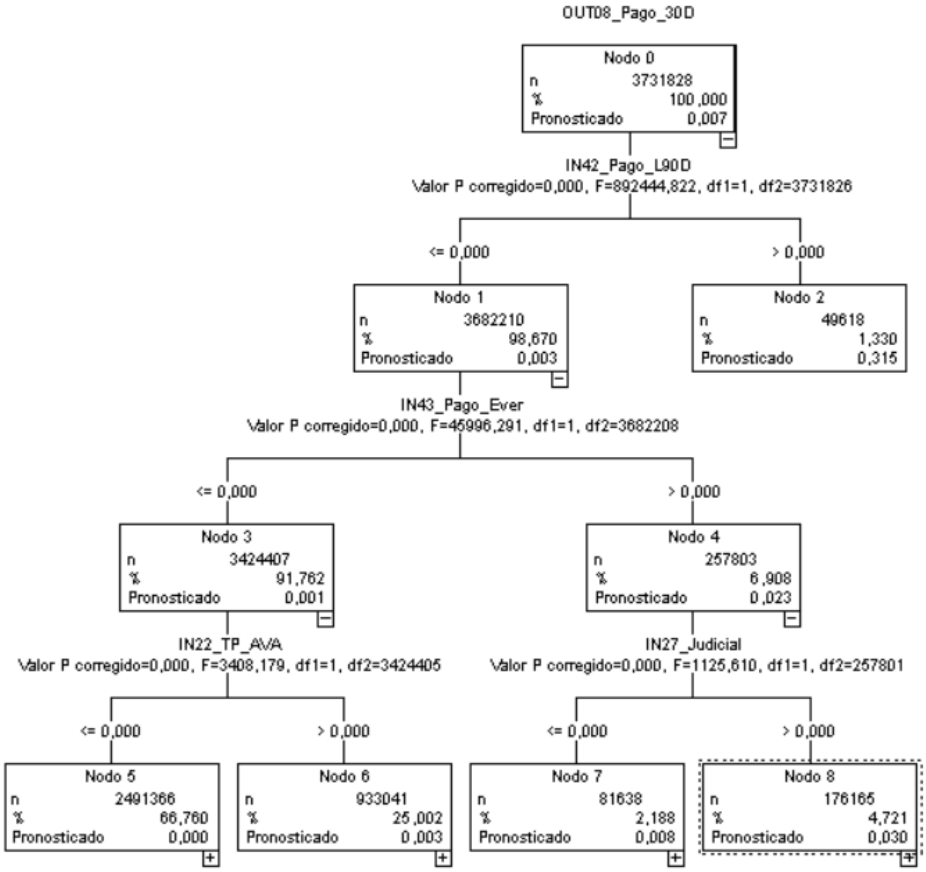


Source: Own elaboration

Model snapshot – initial layers

In the following figure, the initial 3 layers of this model are shown.

Figure 296. Initial layers of the SMEs Post-Covid with payers model (complete)



Source: Own elaboration

Gini and KS indicators and model predictability

In the following table we summarize the Gini and KS indicators for the SMEs Post-Covid model.

Table 297. Gini and K-S indicators of the SMEs Post-Covid with payers model (complete)

Indicators	Training sample	Testing sample
Gini	0,914	0,908
K-S	0,793	0,789

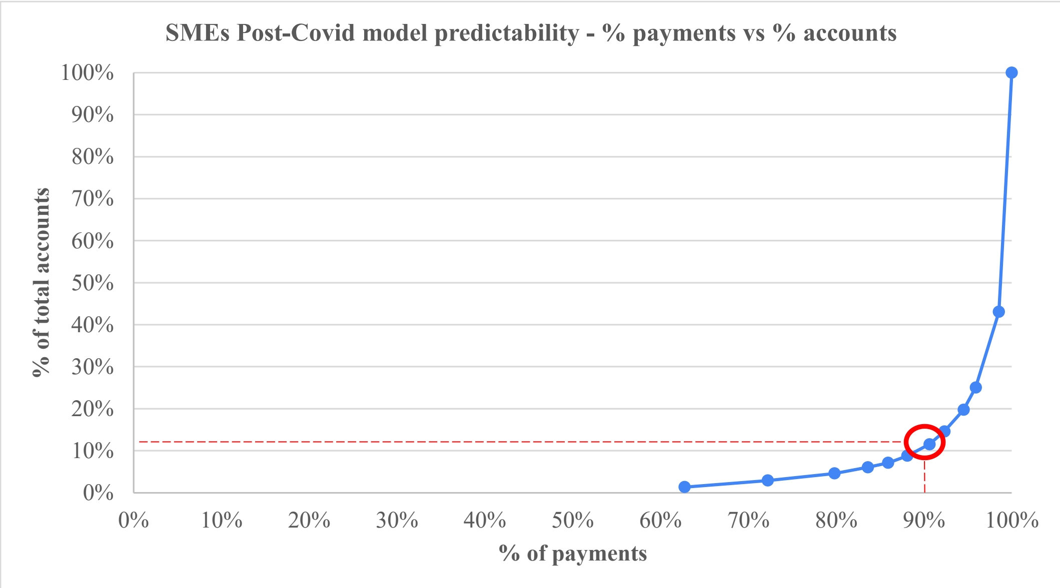
Source: Own elaboration

Model predictability: % of accounts and % of payments

Based on the model built for the SMEs Post-Covid, with a selection of 12% of the accounts, we obtain 91% of the total accounts which will pay in the following 30 days.

Table 298 and Figure 297. Distribution of accounts and payments on the SMEs Post-Covid with payers model (complete)

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	99%	43%
0,002	96%	25%
0,003	95%	20%
0,004	92%	15%
0,006	91%	12%
0,009	88%	9%
0,015	86%	7%
0,017	84%	6%
0,030	80%	5%
0,041	72%	3%
0,315	63%	1%



Source: Own elaboration

APPENDIX 6.- THESIS SUMMARY

DBA
IN MANAGEMENT AND TECHNOLOGY

Universidad Pontificia Comillas

Doctorate of Business Administration in Management and Technology

DEBT SEGMENTATION FOR COLLECTIONS AND RECOVERIES MANAGEMENT IN THE SPANISH MARKET. IMPACT OF A CRISIS: COVID-19 CASE STUDY – SUMMARY

Author: Antón Alfaya González
Director: Dr. Álvaro Caballo Trebol

MADRID | February 2026

COMILLAS UNIVERSIDAD PONTIFICIA ICADE

COMILLAS UNIVERSIDAD PONTIFICIA ICAI

CONTENT INDEX

SECTION 1.- ABSTRACT

Collections and recoveries (C&R) activities are essential for effective credit risk management, particularly in light of post-2007-2008 financial crisis regulations. Despite their importance, this area is often underexplored. Accurate unpaid debt segmentation is crucial for optimizing collection strategies and assessing collection potential. Our objective is to identify the key determinants influencing recovery rates for unsecured debt portfolios that banks and financial institutions have written off, and to assess how these determinants were affected by the COVID-19 crisis. Specifically, we aim to identify which variables significantly predict short-term repayment behaviour before, during, and after the pandemic. To this end, we developed statistical models utilizing the CHAID classification tree methodology to identify the most discriminating variables. Our findings indicate that the COVID-19 crisis had a notable impact, causing significant alterations in these variables. The insights from this research are expected to be valuable to key stakeholders in the C&R sector, including creditors, servicers, and regulators/supervisors. In this research, we utilized a database from a prominent Investment Firm, containing information on 900.000 accounts of unsecured debt portfolios sourced from the Spanish non-performing loans market, and their payment behaviour from 2018 to 2023, along with macroeconomic data from the "Instituto Nacional de Estadística" (INE).

SECTION 2.- INTRODUCTION

2.1- Objective and Research Question

This research aims to enhance understanding of debt segmentation and the impact of crises like the COVID-19 pandemic. It focuses on identifying key factors affecting recovery rates for unsecured debt portfolios that banks and financial institutions have written off and evaluating changes in these determinants due to the pandemic. Specifically, it seeks to determine: which variables predict short-term repayment behaviour before, during, and after the COVID-19 crisis?

2.2- Background and Literature Review

Credit Market and the Relevance of Collections and Recoveries

Credit is widely used across sectors like banking, finance, and retail, offering benefits with deferred payment. However, it carries the risk of non-payment, known as credit risk, which is a long-standing concern in financial markets. The credit risk management cycle includes three key components: Admission, assessing customer eligibility for credit; Portfolio management, establishing risk controls and policies; and Collections & Recoveries (C&R), handling unpaid debts and defaults (Molina et al., 2015).

The 2007-2008 financial crisis significantly increased non-performing assets, leading regulators to strengthen banks' risk management practices. Improvements to the Basel Accords' Internal Ratings-Based (IRB) approach improved credit risk management. The Basel II-III Accords defined key credit risk parameters for estimating expected losses (EL), including probability of default (PD), exposure at default (EAD), loss given default (LGD), and recovery rate (RR). Notably, LGD is calculated as LGD = 1 - RR, where RR indicates the funds recovered from unpaid debts (Carleo & Rocci, 2024; Ramos Gonzalez et al., 2023).

Collections and Recovery Activities

References to unpaid debts and C&R activities have been scarce in scholarly research, but interest in this topic has grown significantly in recent years (Li et al., 2023A).

To recover amounts owed, creditors undertake costly processes, primarily through amicable management, judicial management, and debt sales. Key aspects identified in academic literature include locating debtors, segmentation, and the use of technology (Malinconico & Virguti, 2024; Hunt, 2007).

Research on LGD and RR

Research on LGD and RR reveals two primary streams: the link between PD and LGD-RR, and the estimation methods and key factors influencing both LGD and RR (Han & Jang, 2013).

Main findings show a negative correlation between PD and RR (Altman et al., 2005), and, during recessions, default rates rise while RRs decline (Li et al., 2023A).

A consensus on the best methodology remains unestablished. Historically, regression techniques, particularly ordinary least squares, have been the benchmark (Li et al., 2023A). Decision and regression trees have emerged as strong alternatives to them, often showing better performance (Bastos, 2014).

The literature lacks consensus on the key determinants of LGD and RR, often presenting conflicting findings. Similar to the 5 Cs of credit (Character, Capacity, Capital, Collateral, and Conditions; Rosenberg & Gleit, 1994), our review identifies four main sources of information: borrower characteristics, debt/account characteristics, macroeconomic variables, and repayment behaviour and C&R actions (Xu et al., 2024).

Debt Segmentation and COVID-19 Impact

The adage "past collections do not guarantee future collections" highlights a key principle in the C&R industry. However, analysing past efforts can enhance future outcomes through effective debt segmentation and tailored strategies. This segmentation prioritizes activities, optimizes resource allocation, and improves overall performance across all management phases—preventive, early arrears, late arrears, and written-off (Malinconico & Virguti, 2024; Hunt, 2007).

The COVID-19 crisis, starting in late 2019, caused significant health, social, and economic challenges throughout 2020. Measures to limit virus transmission resulted in a deep economic recession impacting nearly all sectors, including credit risk management and C&R (Ramos Gonzalez et al., 2023).

2.3- Practical Relevance, Research Gap, and Contribution to the Literature

This research analyses debt segmentation in unsecured debt portfolios acquired from banks and financial institutions, highlighting key determinants and the impact of the COVID-19 pandemic. The findings will benefit stakeholders in the C&R

sector, including creditors, servicers, and regulators, by enhancing understanding of how crises affect debt segmentation and improving modelling and outcomes.

While existing literature addresses C&R activities, methodologies, and determinants of RR and LGD, this study uniquely explores effective segmentation variables and COVID-19's effects on debt segmentation for unsecured written-off portfolios acquired by investment firms in the Spanish market.

SECTION 3.- RESEARCH METHODOLOGY

This thesis was developed in five phases, beginning with a literature review of C&R, particularly focusing on RR and LGD modelling. The aim was to identify key debt segmentation determinants and assess effective modelling methodologies. The review aligned academic findings with practitioners' approaches, resulting in a comprehensive list of determinants and methodologies that integrated insights from both sources.

Next, data were gathered from two sources: the Investment Firm systems, for account and debtor characteristics, payment behaviour (2018-2023), and collection activities; and the INE public website for macroeconomic factors. Data cleaning and normalization were conducted to enhance analysis efficiency.

In the third phase, we developed statistical models by initially creating the dependent variable: the probability that an account makes a payment within 30 days. Additionally, we derived additional variables and performed univariate and bivariate analyses to assess their correlation with the dependent variable. The dataset was divided into Individuals and SMEs, using 80% for training and 20% for testing.

Statistical models were developed using CHAID classification trees to predict the defined variable across four periods: Global (2108-2023), Pre-COVID-19 (2018-2019), COVID-19 (2020-2021), and Post-COVID-19 (2022-2023), resulting in eight models.

The models were initially developed with all variables, including clients with prior payments. Due to high correlation with these payments, they were revised to exclude payers. A streamlined version was created with only the most relevant variables, and all models were analysed for predictive quality using key performance indicators like the Gini coefficient and Kolmogorov-Smirnov test.

In the fourth phase, we shared the models with practitioners to gather feedback and compare results with their experiences, leading to adjustments for improvement.

Finally, we analysed and compared models to assess whether the main determinants for segmenting debt portfolios were consistent across Global, Pre-COVID-19, COVID-19, and Post-COVID-19 models, and drew conclusions.

SECTION 4.- RESULTS

4.1- Results Summary and Confirmation of the Research Question

The models developed demonstrated strong predictive capabilities, with Gini coefficients ranging from 0,522 to 0,713 and K-S indicators from 0,379 to 0,577, compared to literature values of 0,7076 for Gini and 0,5647 for K-S (Dumitrescu et al., 2022).

The tables summarize our model variables and their statistical significance, illustrated with a heat map. We also categorized the weights into four types: C&R activities, debt characteristics, debtor characteristics, and macroeconomic factors.

Table 1. Summary of variables weight and "heat" map of the Individuals models

Global		Pre-COVID		COVID		Post-COVID	
Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight
Judicial flag	0,5570	Judicial flag	0,4752	Creditor grouped	0,3230	Creditor grouped	0,5901
RPC I90d	0,2240	Initial other debt amount	0,2476	Judicial flag	0,3202	Judicial flag	0,1824
Creditor grouped	0,1336	Creditor grouped	0,1709	RPC I90d	0,3143	RPC I90d	0,1636
Initial other debt amount	0,0693	National nationality flag	0,0514	Monthly debt amount grouped	0,0239	Product type grouped	0,0201
GDP FTEE annual deciles	0,0116	Youngest DOB	0,0314	Calls ever flag	0,0097	Other nationality flag	0,0192
National nationality flag	0,0034	Region	0,0122	RPC ever flag	0,0050	Calls I90d flag	0,0140
GCPI annual deciles	0,0011	GCPI monthly deciles	0,0075	UCPI monthly deciles	0,0039	UCPI annual	0,0054
GCPI monthly deciles	0,0000	Initial interest debt amount	0,0039			RPC ever flag	0,0032
AR annual deciles	0,0000					GCPI annual	0,0020
						Calls ever flag	0,0000

Global		Pre-COVID		COVID		Post-COVID	
Type of variable	Weight	Type of variable	Weight	Type of variable	Weight	Type of variable	Weight
C&R activities	0,7810	C&R activities	0,4752	C&R activities	0,6492	C&R activities	0,3632
Debt characteristics	0,2029	Debt characteristics	0,4224	Debt characteristics	0,3469	Debt characteristics	0,6102
Debtor characteristics	0,0034	Debtor characteristics	0,0950	Debtor characteristics	0,0000	Debtor characteristics	0,0192
Macroeconomic	0,0127	Macroeconomic	0,0075	Macroeconomic	0,0039	Macroeconomic	0,0074

Source: Own elaboration

Table 2. Summary of variables weight and "heat" map of the SMEs models

Global		Pre-COVID		COVID		Post-COVID	
Variable	**Weight**	**Variable**	**Weight**	**Variable**	**Weight**	**Variable**	**Weight**
Guarantor flag	0,5130	**Guarantor flag**	0,4256	**Guarantor flag**	0,4605	**Guarantor flag**	0,4763
Judicial flag	0,3532	**Judicial flag**	0,3037	**Judicial flag**	0,3594	**Judicial flag**	0,2249
Product type grouped	0,0457	**Monthly debt amount grouped**	0,0895	**Initial principal debt amount**	0,1068	**RPC ever flag**	0,2000
GDP MP annual	0,0233	**Product type grouped**	0,0505	**Creditor grouped**	0,0187	**UCPI annual**	0,0252
Creditor grouped	0,0184	**# phone numbers**	0,0446	**CCI expectations deciles**	0,0160	**Product type grouped**	0,0236
Monthly debt amount grouped	0,0178	**CCI expectations deciles**	0,0220	**CCI**	0,0137	**GCPI monthly**	0,0189
# mobile phone numbers	0,0140	**Calls ever flag**	0,0204	**GDP MP quarter deciles**	0,0082	**Creditor grouped**	0,0166
Initial interest debt amount	0,0112	**Region**	0,0201	**CCI actual deciles**	0,0078	**# fixed line phone numbers**	0,0095
# phone numbers	0,0032	**GCPI monthly**	0,0134	**GDP MP annual deciles**	0,0046	**GCPI annual**	0,0050
		Creditor grouped	0,0088	**Initial interest debt amount**	0,0044	**Monthly debt amount grouped**	0,0000
		CCI actual	0,0014	**UR annual**	0,0000		

Global		Pre-COVID		COVID		Post-COVID	
Type of variable	**Weight**	**Type of variable**	**Weight**	**Type of variable**	**Weight**	**Type of variable**	**Weight**
C&R activities	0,3704	**C&R activities**	0,3687	**C&R activities**	0,3594	**C&R activities**	0,4344
Debt characteristics	0,6061	**Debt characteristics**	0,5744	**Debt characteristics**	0,5904	**Debt characteristics**	0,5165
Debtor characteristics	0,0000	**Debtor characteristics**	0,0201	**Debtor characteristics**	0,0000	**Debtor characteristics**	0,0000
Macroeconomic	0,0233	**Macroeconomic**	0,0368	**Macroeconomic**	0,0503	**Macroeconomic**	0,0491

Source: Own elaboration

By analysing the available data and developed models, we identify the discriminant variables affecting recovery rates. Our findings confirm that these variables have been influenced by COVID-19 in both Individual and SME segments. The pandemic clearly altered the significance of these variables, with a more pronounced effect in the Individuals models.

4.2- General Results of the Research

In analysing Individuals models, notable distinctions appear across time frames. In the Pre-COVID model, the Judicial flag is the most significant variable, followed by Initial other debt amount and Creditor. Variables related to debtor characteristics, such as National nationality flag, Youngest date of birth, and Region, along with

General-Consumer-Price-Index and Initial interest debt amount, are included, though their contributions are marginal.

In contrast, the COVID Individuals model illustrates a shift in variable importance, highlighting three key variables: Creditor, Judicial flag, and Right-Party-Contact last 90 days, which carry similar weights. The Monthly debt amount follows as significant, while Calls ever flag, Right-Party-Contact ever flag, and Underlying-Consumer-Price-Index have marginal importance. This model emphasizes collections-related variables and shows reduced focus on debtor characteristics compared to the Pre-COVID model.

In the Post-COVID Individuals models, the key variables—Creditor, Judicial flag, and Right-Party-Contact last 90 days—remain significant but have shifted in weight. Other variables include Product type, Other nationality flag, and Calls last 90 days; while Underlying-Consumer-Price-Index, Right-Party-Contact ever flag, General-Consumer-Price-Index, and Calls ever flag show marginal significance.

Within SMEs models, the Guarantor flag and the Judicial flag are key variables, though their weights vary across models. Moreover, the four variable categories—debt characteristics, debtor characteristics, collection activities, and macroeconomic factors—vary in significance.

In the Pre-COVID SMEs model, key variables include the Guarantor flag, Judicial flag, and Monthly debt amount. Additional variables are Product type, Number of phone numbers, Consumer-Confidence-Index expectations, Calls ever flag, Region, and General-Consumer-Price-Index, with Creditor and Consumer-Confidence-Index actual showing marginal significance.

Transitioning to the SMEs COVID model reveals a significant shift in variable importance. The Guarantor flag, Judicial flag, and Initial principal debt amount are the top three variables, with greater weight than in the Pre-COVID model. Subsequent variables, like Creditor and Consumer-Confidence-Index expectations, receive lower weights. Notably, the weight of collection activities has declined compared to the Pre-COVID model.

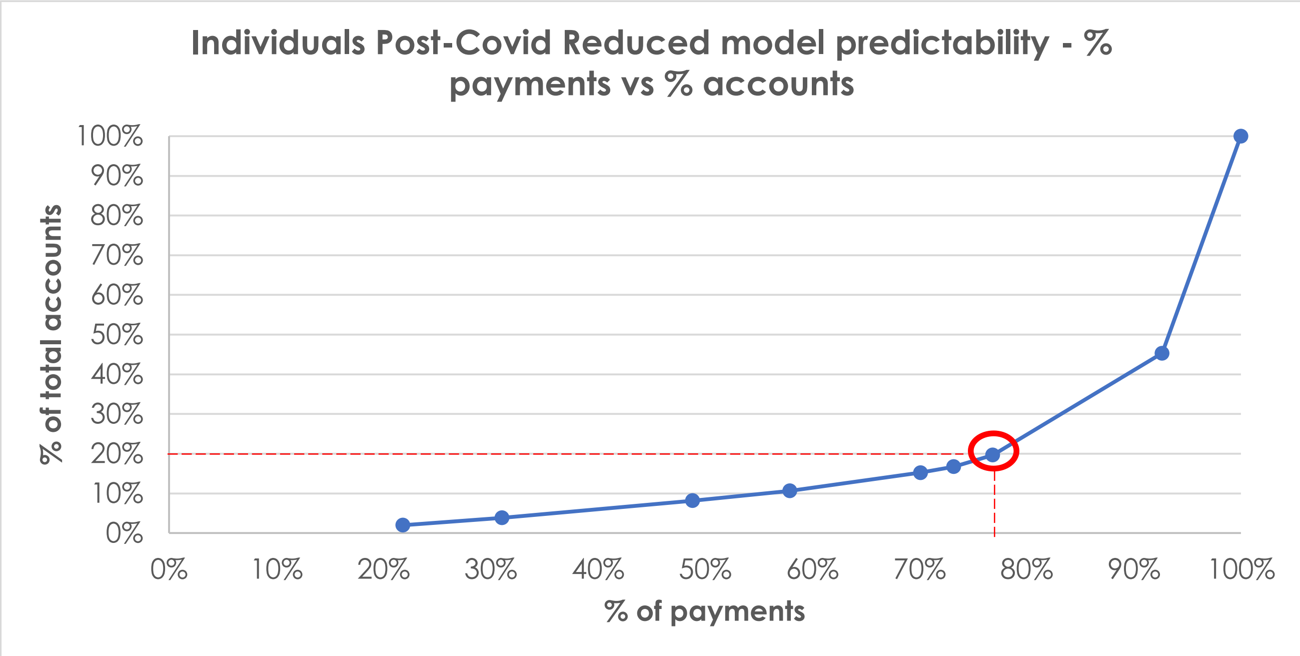
Lastly, the SMEs Post-COVID model emphasizes the Guarantor flag, Judicial flag, and Right-Party-Contact ever flag as primary variables, alongside factors like Underlying-Consumer-Price-Index annual, Product type, monthly General-Consumer-Price-Index, and Creditor. It also considers variables such as the Number of fixed line numbers, annual General-Consumer-Price-Index, and Monthly debt amount, though they hold marginal weight.

4.3- Illustrative Application of the Models Built

Segmentation models in C&R are crucial for resource optimisation. For illustration, our Individuals Post-COVID model, with a 20% sample, predicts 77% of the total accounts that will pay in the following 30 days, highlighting the model's effectiveness in improving recovery strategies.

Table 3 and Figure 1. Distribution of accounts and payments on the Individuals Post-COVID model

MODEL SCORE	% PAYMENTS	% ACCOUNTS
0,000	100%	100%
0,001	93%	45%
0,002	77%	20%
0,003	73%	17%
0,004	70%	15%
0,005	58%	11%
0,006	49%	8%
0,007	31%	4%
0,016	22%	2%



Source: Own elaboration

SECTION 5.- CONCLUSIONS

5.1- Discussion of the Research Results

The primary sources for LGD-RR modelling include borrower characteristics, debt characteristics, macroeconomic variables, and C&R activities. Recent literature (Xu et al., 2024) and our research confirm their relevance for both Individuals and SMEs segments, though highlighting notable differences between the segments and earlier studies.

In the literature, borrower and debt characteristics are frequently discussed, macroeconomic factors occasionally, and some studies recently include repayment behaviour and C&R variables (Xu et al., 2024; Li et al., 2023A). Our models emphasize debt characteristics and C&R variables for both segments, diverging from the traditional focus on borrower and debt characteristics.

Borrower characteristics are minimal features in our models. In the SMEs model, only Region appears in the Pre-COVID version. The Individuals model includes Nationality, Age, and Region in the Pre-COVID version, but these characteristics are absent in the COVID model, and have limited presence in the Post-COVID and Global models.

Integrating C&R activities has improved our modelling, aligning findings from previous authors (Bellotti et al., 2021; Han & Jang, 2013), which show that including debt collection activities in LGD modelling enhances predictive accuracy over models using only firm-specific variables.

Our conclusions differ from much of the existing literature mainly due to our focus on written-off portfolios with an average debt age of 11-12 years, compared to the younger debt profiles typically examined in other studies, and the limited access to information on C&R activities.

In RR modelling, borrower characteristics are more significant for newer debts, while C&R activities gain importance for older debts, reducing the impact of borrower traits. Access to repayment behaviour and C&R data is typically limited, but having this information significantly enhanced our modelling efforts.

In C&R practices, incorporating macroeconomic variables into RR modelling is less common, yet our research shows they significantly enhance predictive power, consistently appearing in our final models. However, as seen in previous literature (Han & Jang, 2013), their impact is reduced.

Our models incorporate key macroeconomic variables from established academic references, including the Consumer-Price-Index, GDP, Consumer-Confidence-Index, and activity/unemployment rates (Xu et al., 2024; Galow et al., 2024; Xia et al., 2021).

5.2- Recommendations Based on the Research Results

This research findings are valuable academically and for key stakeholders in the C&R sector: creditors, servicers, regulators, and supervisors.

For creditors and servicers managing debt portfolios, modelling potential RR enhances efficiency and effectiveness in portfolio management and valuation. It allows targeted collection efforts on accounts with higher recovery likelihood, reducing wasted resources on less collectible accounts. Additionally, it improves understanding of potential recoveries, enhancing pricing accuracy for portfolio transactions.

Incorporating debt characteristics, debtor traits, C&R activities, and macroeconomic factors in modelling is essential, and remarkably, in our research, debt characteristics and C&R activities are the key sources. Focus on variables like the Judicial flag, Creditor, and Right-Party-Contact for Individual accounts, and the Judicial flag, Guarantor flag, and Right-Party-Contact for SMEs. Additionally, entities should enhance their RR segmentation models, especially in response to significant events like the COVID-19 pandemic, which significantly affect Individuals accounts more than SMEs.

This research emphasizes the significance of segmentation models in C&R for regulators and supervisors. These models enhance C&R efficacy and efficiency by considering the key sources highlighted, ultimately supporting a stronger economy. The COVID-19 pandemic has influenced RR determinants, indicating that best practices in segmentation must adapt to future challenges to strengthen creditor resilience during downturns.

SECTION 6.- LIMITATIONS AND STRENGTHS

This research has limitations, including its focus on unsecured debts that are 11-12 years old. Incorporating other debt types, such as secured or more recent debts, could have strengthened the analysis. Additionally, the lack of comprehensive data, including judicial record details, credit bureau and scoring information, and SMEs financial insights, hindered our modelling efforts.

In contrast, the developed models exhibit notable strengths. They demonstrate strong performance and align well with existing literature, showcasing their robustness, as the most pertinent variables remain stable over time, with fluctuations primarily observed in their weights. These models were supported by a comprehensive dataset of approximately 900.000 accounts monitored over a six-year period (2018-2023), which includes borrower characteristics, debt attributes, macroeconomic factors, and repayment behaviour, along with C&R activities. Furthermore, this study represents the first investigation into effective segmentation variables and the impact of COVID-19 on debt segmentation for C&R of unsecured written-off portfolios acquired by investment firms in the Spanish market.

SECTION 7.- REFERENCES

Altman, E. I., Brady, B., Resti, A., & Sironi, A. (2005). The link between default and recovery rates: Theory, empirical evidence, and implications. *The Journal of Business*, *78*(6), 2203-2228.

Bastos, J. A. (2014). Ensemble predictions of recovery rates. *Journal of Financial Services Research*, *46*(2), 177-193.

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

Carleo, A., & Rocci, R. (2024). Functional clustering of NPLs recovery curves. *Socio-Economic Planning Sciences*, *95*, 102018.

Dumitrescu, E., Hué, S., Hurlin, C., & Tokpavi, S. (2022). Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *European Journal of Operational Research*, *297*(3), 1178-1192.

Galow, B., Georgescu, O. M., & Marques, A. P. (2024). *Loss-given-default and macroeconomic conditions* (No. 2954). ECB Working Paper.

Han, C., & Jang, Y. (2013). Effects of debt collection practices on loss given default. *Journal of Banking & Finance*, *37*(1), 21-31.

Hunt, R. M. (2007). Collecting consumer debt in America. *Available at SSRN 993249*.

Li, A., Li, Z., & Bellotti, A. (2023A). Predicting loss given default of unsecured consumer loans with time-varying survival scores. *Pacific-Basin Finance Journal*, *78*, 101949.

Malinconico, A., & Virguti, E. (2024). Business Models and Best Practices in Debt Recovery Services: an overview of the Italian industry. In *Conference Proceedings*.

Molina, S. (2015). *El ciclo del crédito*. Editorial Almuzara.

Ramos González, M. R., Ureña, A. P., & Fernández-Aguado, P. G. (2023). Forecasting for regulatory credit loss derived from the COVID-19 pandemic: A machine learning approach. *Research in International Business and Finance*, *64*, 101907.

Rosenberg, E., & Gleit, A. (1994). Quantitative methods in credit management: a survey. *Operations research*, *42*(4), 589-613.

Xia, Y., Zhao, J., He, L., Li, Y., & Yang, X. (2021). Forecasting loss given default for peer-to-peer loans via heterogeneous stacking ensemble approach. *International Journal of Forecasting*, *37*(4), 1590-1613.

Xu, Q., Liu, C., Luo, J., & Liu, F. (2024). Using machine learning to investigate the determinants of loan default in P2P lending: Are there differences between before and during COVID-19?. *Pacific-Basin Finance Journal*, *88*, 102550.