MASTER

MATHEMATICAL FINANCE

MASTER´S FINAL WORK

INTERNSHIP REPORT
(PROVISIONAL DOCUMENT)

PROBABILITY OF DEFAULT: MODELLING AND BACKTESTING

RAFAEL FERREIRA GRANGEIA

SUPERVISION:
RAQUEL LOURENÇO
ONOFRE SIMÕES

OCTOBER - 2024

ACRONYMS

BCBS – Basel Committee on Bank Supervision

BNPP – BNP Paribas

DR – Default Rate

EAD – Exposure at Default

EBA – European Bank Authority

ECB – European Central Bank

GEE - Generalized Estimating Equations

HQLA – High-Quality Assets

IRB - Internal Ratings-Based

IRBA - Internal Ratings-Based Advanced

IV – Informative Value

KPI – Key Performance Indicators

LCR – Liquidity Coverage Ratio

LGD – Loss Given Default

MoC – Margin of Conservatism

PD – Probability of Default

PiT – Point in Time

RC – Risk Class

RDS – Reference Data Set

RWA – Risk Weighted Assets

TTC – Through the Cicle

WoE – Weight of Evidence

ABSTRACT

Basel III introduced the Internal Rating Based (IRB) and IRB-Advanced (IRBA) approaches, which allow banks to use their own internal estimates of risk parameters to calculate the necessary regulatory capital requirements for credit risk. While the IRB approach enable banks to create and utilize sophisticated risk models adapted to their unique experiences and data, the IRBA methodology grants banks even greater discretion, allowing them to estimate all risk components independently, provided they meet specific criteria and obtain regulatory approval.

Backtesting is a crucial process in financial risk management, employed to assess the performance and reliability of models over time. This practice is essential for maintaining robust risk management systems and ensuring compliance with regulatory requirements. By comparing predicted risk estimates with actual outcomes, backtesting helps in identifying discrepancies, ensuring that models remain accurate and relevant under changing market conditions.

The Probability of Default (PD) parameter is a risk input that measures the likelihood that a borrower will default on their debt obligations in a specific date. This report focuses on the development of a PD model and its subsequent validation through Backtesting, ensuring its alignment with regulatory standards.

The PD model development followed a structured approach, utilizing logistic regression combined with K-means clustering to form distinct risk classes, each assigned a specific PD. A scoring system was designed to rank obligors by risk, incorporating the Margin of Conservatism (MoC) to provide a buffer against potential risk underestimations, thereby enhancing model reliability.

The backtesting framework was evaluated on four dimensions: stability, discriminatory power, calibration accuracy, and conservatism. Three scenarios were simulated to test the model's robustness.

Results indicated that the PD model generally maintained stability and discriminatory power, though calibration issues and heterogeneity in clusters were observed. The model was conservative, overestimating risk.

Keywords: Credit Risk, Probability of Default, Backtesting, Cluster, Dimensions

TABLE OF CONTENTS

LIST OF TABLES AND GRAPH

ACKNOWLEDGEMENTS

I could not begin without thanking my family, especially my mother Rosa, my brother Tiago, my sister-in-law Denise, and my two incredible nephews, Tomás and Duarte. Their efforts made it possible for me to come and study in Lisbon, embarking on a journey that started in 2019. I am deeply thankful for their continued motivation throughout this work and my entire academic journey, for always supporting my decisions and being there for me.

Secondly, my appreciation goes to my colleagues at the university. From the moment I entered ISEG, no one ever refused to help me, even when I might have bored them. Thank you very much for making these last five years the best years of my life.

To my friends, for being by my side, providing endless laughter, and promising a celebration upon the completion of this work.

I am grateful to all my professors at ISEG for providing the foundation to develop this work and for always being available to help.

I would like to extend my thanks to everyone at BNPP, especially my team for making office days enjoyable and for the warm welcome from the very beginning. My supervisor, Raquel Lourenço, for all the times she inquired about my work and found time to advise me. To Fátima, Béchir, David, Ignácio, and Antoine who, in various ways, helped me during this work.

There is a coworker who deserves a special mention. Steven, I have no idea how to repay all the effort, help, and advice you have given me over the last eight months. Thank you very much for all the time you spent helping me, often after work. I do not know what I could have done without your assistance. I will certainly keep my promise and pay you a well-deserved lunch.

Finally, a special word of thanks to my ISEG supervisor, Professor Onofre Simões, for his invaluable assistance, ideas, and guidance throughout this process. You will always have a place in my heart, not only for your help with this work but also for all our conversations and classes over the past five years. I am truly grateful to you.

1. INTRODUCTION

The present work is the result of an 8 - month curricular internship at BNPP – Banque Nationale de Paris et Paribas, in the RISK Models & Regulatory department, with the purpose of concluding the Master program in Mathematical Finance, at ISEG − Lisbon School of Economics and Management. Being a member of the Model Performance team, I worked on Credit Risk, more specifically assessing and validating through the Backtesting exercise the performance of the internal model, estimating the Probability of Default (PD) risk parameter in Mid Corporate scope[1].

Backtesting is a key process in financial risk management, used to tracking the performance and reliability of the used models over time. This practice is essential for maintaining robust risk management frameworks and complying with regulatory requirements. Internal Ratings-Based (IRB) Models are a fundamental element of banking risk management, especially under the Basel II and Basel III regulatory frameworks. By using IRB models, banks can calculate their capital requirements for credit risk based on internal estimates of risk components such as PD, Loss Given Default (LGD), Exposure at Default (EAD). In the IRB context, Backtesting is crucial to ensure that these internal models have predict risk accurately.

During my time at BNPP, I have conducted a Backtesting exercise, which primarily involved two tasks: constructing the Reference Data Set (RDS) and conducting Backtesting analysis using internal tools to assess the model's performance. Accordingly, this work will focus on the development of a model to estimate PD and its subsequent Backtesting.

To provide context, this report is structured as follows: First a presentation of the IRBA regulation history and current practice followed by an explanation of the general overview of PD models and how they are usually constructed in the context of IRBA. Based on this, I will explain how I have created from scratch a PD Model based on BNPP open source data and simulations. Following this, a presentation of the backtesting tests aiming at ensuring the perennity of PD models. Finally, I will simulate 3 scenarios to observe the reaction of the model and the capacity of the backtesting tests to highlight possible issues.

[1] This is an intermediary scope between small companies and Large Corporate companies. Usually, the turnover concerns a range of values that go from few hundred thousand to a few dozen millions of euros.

2. BASEL ACCORDS

The Basel Accords are a comprehensive set of standards established by the Basel Committee on Banking Supervision (BCBS), the leading global authority on prudential regulation for banks. BCBS members have committed to fully implement these standards and ensuring their application by internationally active banks within their respective jurisdictions.

2.1 Basel I

In 1988, the Basel I Accord was published, requiring internationally active banks to maintain a minimum capital requirement of 8% of their risk-weighted assets (RWA). The main aim of the Accord was to minimise credit risk, establishing how much capital should financial institutions keep in reserve to guarantee they would undertake their obligations. It was first introduced in the BCBS countries and is now applied in all countries where internationally active banks operate, helping these countries to standardize their rules.

In addition, Basel I aims to establish a sufficient level of capital adequacy, which refers to the risk of unexpected losses that have a negative impact on these institutions. It subdivides assets into five different risk categories, defined as follow: 0% (cash, central bank debt, government debt and all Organisation for Economic Cooperation and Development (OECD) debt), 10% (central bank debt of countries with a high percentage of inflation), 20% (bank debt, bank development debt, non-OECD public sector debt and non-OECD bank debt with a maturity of over one year), 50% (residential mortgages) and 100% (private sector debt, capital instruments issued by other banks, real estate, plant and equipment and non-OECD bank debt with a maturity of over one year). The previous approach does not differentiate between potential discrepancies in the creditworthiness of each individual borrower within each category.

RWA is a measure used to calculate the minimum amount of capital that a financial institution must hold according to the risk profile of its assets. This capital is then allocated to cover unexpected losses that may occur. By assessing their RWA, financial institutions get a better understanding of the risks underlaying that each asset apports and how they should allocate capital accordingly.

Over time, however, the Basel I Accord became unable to address adequately the increasing complexity and associated risks of the banking industry. Consequently, a

revised framework was devised to more accurately align regulatory capital with the underlying risks faced by international banks.

2.2 Basel II

In June 2004 Basel II is launched as the result of several purposes and revisions of the previous Accord. The new Accord is ordered according to three fundamental pillars. Under Pillar I, the new framework sets out criteria to develop and expand the standardised rules for banking organizations to adopt more risk-sensitive minimum capital requirements. It lays out principles for banks to assess the adequacy of their capital, introducing new risk-sensitive options for the computation of credit risk (standardised approach, foundation internal ratings-based approach, advanced internal ratings-based approach) and operational risk. Pillar II outlines the principles that supervisors are to employ in reviewing the assessment of capital adequacy, with the objective of ensuring that banks have adequate capital to support their risks. Pillar III provides for the enhancement of market discipline by requiring that investors be provided with all relevant information necessary to assess the risk profile of a bank.[2]

The main innovations introduced by Basel II were important in defining risk assessment strategies and helped to increase banks' transparency, but they did not take into account an aspect that came to the fore in 2007-08 with the US financial crisis: liquidity risk. This became evident during that period when many financial institutions struggled to access sufficient liquidity, thereby exacerbating the crisis.

2.3 Basel III

In December 2010, the BCBS presented Basel III as the result of an effort to strengthen the regulation and supervision of internationally active banks in the light of the weaknesses revealed by the financial crisis. In terms of specific measures, the changes and innovations stand out in 3 main areas:

i) Definition of capital and minimum capital requirements: The definition of capital has been subject to a process of harmonisation between jurisdictions and it has become mandatory to verify minimum capital requirements for 3 levels of capital quality, Common Equity Tier 1 Ratio; Tier 1 Ratio: minimum; Total Capital Ratio.

[2] The Comprehensive Approach of Basel II (europa.eu)

ii) Capital buffers (which act as a cushion in times of economic stress and ensure that banks have sufficient capital to absorb losses) and anti-procyclicality measures:

iii) Leverage and liquidity measures: Basel III introduced new leverage and liquidity requirements to safeguard against risky lending while ensuring banks maintain adequate liquidity during financial stress. These measures include a higher leverage ratio for G-SIBs and new liquidity rules such as the Liquidity Coverage Ratio (LCR).

Basel III expands the scope of risk management beyond credit, market and operational risks to include liquidity risks. The inclusion of liquidity standards and leverage ratios enhances the banking sector's resilience to a broader range of risks, fostering a more stable financial environment. By introducing the G-SIB buffer, Basel III takes a comprehensive approach to addressing systemic risk, a significant advancement from Basel II.

2.4 IRB and IRBA Models

A key component of Basel III is the development of the IRB and IRB-Advanced (IRBA) approaches, which enable banks to use their own internal estimates of risk parameters to calculate the necessary regulatory capital requirements for credit risk.

The IRB approach allows banks to create and utilize sophisticated risk models tailored to their unique experiences and data, thereby aligning regulatory capital more closely with actual risk. This flexibility incentivizes banks to continuously enhance their risk management practices. However, the use of IRB models requires prior approval from the relevant regulatory authorities, who also monitor the models' accuracy and reliability.

The IRB Advanced approach grants banks even greater discretion, allowing them to estimate all risk components independently, provided they meet specific criteria and obtain regulatory approval. This approach demands the use of advanced data and modelling techniques, as well as the establishment of comprehensive governance and risk management frameworks to support models.

The desire from banks to use IRBA models is mainly driven by for more accurate and risk-sensitive measures of credit risk. Indeed, thanks to the statistical model underlying, the resulting RWA is closely fitting each asset’s risk while the other possible approaches (Foundation and Standard) are purposely overestimating it. Consequently, for a given amount of Own Founds, a bank should be able to have much more assets by using IRBA approach to compute its RWA. Another argument in favour of IRBA is the will from banks to have a precise vison of their portfolio’s risk.

3. SETTING AND BACKTESTING A PD MODEL

3.1 Modelling PD

Bandyopadhyay (2016), states that the PD quantifies the likelihood that borrowers will fail to meet their contractual obligations leading them to default, usually with an horizon of 1 year. While default does not always result in immediate losses, it is clearly increasing the risk of potential financial losses. One common approach when constructing such models is to score borrowers based on their risk profiles. After the scoring phase, clients with a similar risk profile (similar score) are grouped together to Risk Classes. Finally, a PD is determined for each of these classes during the calibration phase. A PD model is therefore a model that consists in a set of scoring, clustering and calibration techniques.

Risk scoring models are a key tool to credit risk management, helping financial institutions assess the creditworthiness of clients and the likelihood of loan non-repayment. These models assign a score to each client, reflecting the risk they pose. The score is derived by developing a scoring function based on the premise that past observations can predict future client behaviour. By utilizing a database containing available client information, a risk criterion is selected to represent the risk level, modelled using various explanatory variables from the database. The goal of the scoring function is to assign a low score to low-risk individuals and a high score to those with higher risk levels.

For modelling the PD, logistic regression models are usually used. Logistic regression models are well-suited for both continuous and categorical variables. These models offer substantial interpretative power by transforming values across the entire real number range into values between 0 and 1, thus predicting the probability of a binary outcome. The logistic regression consists in predicting the outcome of a binary random variable Y - that takes the value 0 (performing) or the value 1 (default) – using explanatory variables denoted X, see Ranganathan et al. (2017).

Considering a sample of observations which is described by a set of the p explanatory variables $X = (X_1, \dots, X_P)$, let $\{X_i = (X_1, \dots, X_P), Y_i\}_{i=1}^{n}$ be the observation associated with the individual i of being risky (the default event $Y_i = 1$) given his behaviour described by X_i. Let π_i be this probability.

$$\pi_i = \mathbb{P}(Y_i = 1|X_i) \quad (1)$$

Ranganathan et al. (2017) also state that, when modelling PD, a common approach consists in using the Generalized Linear Models (GLM). The modelling of the variable Y will be conducted using a bijective function g called link function and expressed as follows:

$$g(\mathbb{P}(Y = 1|X = x) = x\beta \quad (2)$$

where $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$ is a vector of the model's parameters.

The logistic regression is a special case of GLM based on the hypothesis that the random variable, i.e., Y_i, follows a Bernoulli distribution. The link function, called logit function, is the inverse of the logistic cumulative distribution function:

$$g: [0,1] \to \mathbb{R} \quad (3)$$

$$p \to log\left(\frac{p}{1-p}\right) \quad (4)$$

Therefore, for an individual i, the logistic model can be expressed as follows:

$$log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \sum_{j=1}^{p} \beta_i X_{ij} \quad (5)$$

return, the probability of an individual i being risky given their behaviour can be obtained as below:

$$\pi_i = \frac{1}{1 + e^{-\left(\beta_0 + \sum_{j=1}^{p} \beta_i X_{ij}\right)}} \quad (6)$$

To estimate this probability, first we must estimate the coefficients or parameters $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$ of the model. Parameters' estimation is conducted with the maximum likelihood method as it being assumed that Y_i follows as a Bernoulli distribution, which implies the following relation:

$$\forall_i \in [0,1], \mathbb{P}(Y_i|X_i) = {\pi_i}^{Y_i}(1 - \pi_i)^{1-Y_i} \quad (7)$$

The likelihood function $\mathcal{L}(\beta)$ depends on the parameters of the model given a sample of observations $(X_i, Y_i)_{i=1}^n$. The likelihood function associated with the logistic regression is given by:

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} \mathbb{P}(Y_i|X_i) \quad (8)$$

The logarithm of the likelihood function is more convenient to work with and can e deduced from the expression above:

$$log\ \mathcal{L}(\beta) = \sum_{i=1}^{n} Y_i log\pi_i + (1 - Y_i)\log\,(1 - \pi_i) \quad (9)$$

The solution of the maximization of log-likelihood function is unique and corresponds to the maximum likelihood estimator.

The main goal of scoring is to rank obligors with respect to their risk of default and then define score value intervals to construct the Risk Classes to which a PD will be associated.

For determining these score's intervals, the K-means clustering algorithm can be used. This algorithm is an unsupervised method because it begins without labels and then forms and labels groups on its own, see Ikotun et al. (2022). The process begins with randomly assigning each data point to an initial group and calculating the centroid for each one - a centroid is the centre of the group (usually defined as the average of that group).

The method then proceeds to assess each observation, categorising it according to the cluster it is most closely aligned with (the definition of "closely" is that the Euclidean distance between a data point and a group's centroid is shorter than the distances to the other centroids). When a cluster gains or loses a data point during this process, the centroid is recalculated. The procedure continues iteratively until the algorithm is unable to find a more optimal grouping.

Upon completion of the K-means clustering algorithm, all groups have achieved the minimum within-cluster variance, which ensures that they remain as compact as possible. Sets with minimum variance and size exhibit data points that are as similar as possible.

Ensure that the clusters effectively differentiate risk among the obligors is a requirement of the regulation "*Analyses of discriminatory power for PD models should be designed to ensure that the ranking of obligors/facilities resulting from the rating methodology appropriately separates riskier and less risky obligors/facilities.*"[3]. To do so, it is essential to test the heterogeneity among the Risk Classes. This is achieved by conducting a Welch Test. Welch's Test[4] is a statistical test used to determine whether the means of two groups are significantly different from each other. Compared to Student's t-Test, it can handle in a better way situations where the two groups have unequal variances and different sample sizes. The hypotheses are:

[3] Article 65, page 27 of ECB guidelines: ECB guide to internal models (europa.eu)
[4] Welch's t-test: When to Use it + Examples (statology.org)

$\mathbf{H_0}: \mu_1 = \mu_2$ (no difference in the means of the two clusters)

$\mathbf{H_1}: \mu_1 \neq \mu_2$ (the means of the two clusters are different)

where

$$t_{statistic} = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (10)$$

where $\overline{X_i}$, s_i *and* n_i are, respectively, the mean, standard deviation and number of obligors of the Risk Class, i = 1,2.

The final step is to calibrate a PD for each Risk Class. The natural choice is to calibrate the PD estimate of each Risk Class on the Long Run Average (LRA) DR - is the observed average DR - which is calculated as the arithmetic mean of the one-year DRs, across the observed Risk Class.

$$DR_{RC_i}^{observed} = \frac{\sum_k DR_{RC_i}^k}{\sum_k 1} \quad (11)$$

where k and i refers, respectively, to each reference date and Risk Class considered.

To be noted that, during the model's life, a recalibration of the model can be required (for example, due to bug discrepancies between estimation and observations) which leads to an extension of this sample in order to include more recent data.

Inaccuracies, uncertainties or gaps in information regarding the business, as well as methodological or statistical approaches and historical data, can distort the quantification of risk parameters. To address these issues, the ECB requires institutions to incorporate a Margin of Conservatism (MoC) into their estimates, which accounts for the expected range of estimation errors: "*[…] If an institution cannot provide sufficient proof that the external data are representative, in the ECB's view it may still use external data if it shows that the information gained from the use of the external data outweighs any drawbacks stemming from the deficiencies identified and an appropriate margin of conservatism (MoC) is applied*."[5]. This MoC should reflect uncertainty at the level of the final risk estimates, whether at the class or model level. Additionally, institutions must maintain monotonicity in their final estimates while still accounting for this uncertainty.

[5] Article 36, page 67 of ECB guide to internal models (europa.eu)

3.2 Backtesting the PD Model

To effectively manage model risk, financial institutions must establish validation processes that continuously monitor model quality. Backtesting is a crucial quantitative validation tool in this process. It involves comparing the predicted risk measure (PD) with new observations outcomes using a range of statistical tests to evaluate the model's calibration, discrimination power, and stability. Early detection of underperformance is vital, as it directly influences profitability and risk management strategies. For example, in case the model shows bad backtesting results, the regulator has a large variety of possible actions such as a RWA penalty, ask for a revision of the model or even ban the model (leading the bank to use Foundation or Standard approach for the concerned portfolio).

3.2.1 Backtesting Evolution

Banking Backtesting has developed significantly over the past few decades, emerging as a crucial component of financial risk management and regulatory oversight.

Backtesting in banking began in the 1970s and 1980s alongside advancements in quantitative finance and risk management. Financial institutions started using statistical models to predict risks like market risk and credit risk. However, the formal practice of Backtesting these models for predictive accuracy was not yet established, with the focus more on model development than on rigorous validation.

The Basel Accords, as already mentioned, marked a major step in banking Backtesting. Basel I focused on credit risk, setting the stage for more advanced risk management. Basel II explicitly required banks to validate and backtest their internal risk models using historical data, making Backtesting a regulatory necessity. Basel III developed post financial crisis, emphasized the need for continuous Backtesting to ensure model accuracy and reliability under varying market conditions.

Nowadays, advances in technology have led to more automated Backtesting processes, leveraging data analytics, machine learning and big data. Regulatory expectations have also evolved, with enhancements to the Basel framework and the addition of stress testing and scenario analysis as complementary to Backtesting.

3.2.2 Dimensions of Backtesting

According to Castermans et al. (2010), Backtesting involves essentially 4 dimensions: Stability, Discriminatory Power, Calibration accuracy and Conservatism of the model. These four characteristics are critical for detecting potential issues early, helping to maintain the model's relevance and effectiveness, ensuring thus a robust Backtesting procedure.

Stability

Stability assesses how closely the characteristics of the current population align with the population used to develop the model. Over time, changes in bank strategies or population shifts can lead to differences that may cause the model to become less effective.

One indicator that allows to measure this variation is the Population Stability Index (PSI)[6]. It can be computed as follows.

$$\text{PSI} = \sum_{i=1}^{n} (N_i^1 - N_i^0) \times ln\left(\frac{N_i^1}{N_i^0}\right) \tag{12}$$

where N_i^j represents the percentage of population observed in class *i* at time t_j, $j = \{0,1\}$.

Discriminatory Power

Discrimination evaluates the model's ability to distinguish between obligors based on their risk profiles. For a PD model, this means accurately assigning defaulted obligors to higher-risk clusters and non-defaulted obligors to lower-risk clusters.

An indicator that grants one to access the discriminative power of a model is the Accuracy Ratio (AR). Open Risk Academy (2024) defined the AR as the ratio of the area under the Cumulative Accuracy Profile (CAP) of the model under consideration and above of the perfect discriminating model. The CAP is a graphical representation used to assess the discriminatory power of a rating model. It plots the cumulative proportion of actual defaults against the cumulative proportion of obligors, ranked by their predicted risk scores. In a perfect model, all actual defaults would be captured at the beginning of the ranking, resulting in a steep initial curve. The area under the CAP curve indicates how well the model discriminates between defaulting and non-defaulting obligors. The AR

[6] Population Stability Index and Characteristic Analysis (listendata.com) and BNPP internal documentation.

can take values between zero and one and the closer AR is to one, the higher the discriminative power of the classification.

$$AR = \frac{a_R}{a_P} = \frac{2\int_0^1 CAP(u)\,\partial u - 1}{1 - p} \tag{13}$$

where a_R and a_P represent, respectively, the area for the actual and perfect model. To compute area below the current model it was used the trapezoidal rule numerical integration method.

Calibration Accuracy

Calibration measures the gap between the model's predicted outcomes and actual observations. A well-calibrated model is one where the estimated PD closely matches the observed DR, ensuring the model remains reliable over time. The "acceptable distance" between the PD estimated and the observed DR depends directly on the size of the population and the confidence interval chosen. Indeed, the more obligors and the more defaults are observable, the smaller this "acceptable distance" should statistically be.

To assess Calibration accuracy, a Binomial Test should be performed. As stated by Soch et al. (2024), the Binomial Test is a statistical test used to determine whether the estimated parameter is close to the observed one.

Consider a sequence of *n* obligors, each represented by a variable X_i that follows a Bernoulli distribution, where:

- $X_i = 1$ if the obligor defaults
- $X_i = 0$ if the obligor does not default.

It is crucial to note that these variables $X_1, X_2, \ldots, X_n$ are independent and identically distributed (iid). The sum of these variables, $S_n = X_1 + X_2 + \cdots + X_n$, represents the total number of defaults and follows a Binomial distribution $S_n \sim Binomial\ (n, p)$, where p is the probability of success (i.e., the obligor being in default). The Binomial Test involves the following hypotheses:

$\mathbf{H_0}$: The PD estimated equals the LRA DR.

$\mathbf{H_1}$: The PD estimated does not equals the LRA DR.

By using the Binomial Test and leveraging the Central Limit Theorem, the Calibration accuracy of the model can be effectively assessed. Ensuring the p-value is above the

chosen significance level indicates a well-calibrated model, while a p-value below the significance level suggests the need for model adjustments.

If defaults are not iid, a different strategy must be employed. One approach is to use Generalized Estimating Equations (GEE), which can account for correlations within groups of obligors or time-series dependencies. Ming (2014) refers that these models have the ability to provide unbiased estimates of population-averaged regression coefficients, even when the correlation structure is not correctly specified.

Conservatism

For regulatory purposes, this can be assessed by conducting a one-sided binomial test. The one-sided binomial test involves the following hypotheses:

$\mathbf{H_0}$: PD (incorporating the MoC) $\geq$ LRA DR (PD is not underestimated)

$\mathbf{H_1}$: PD (incorporating the MoC) $<$ LRA DR (PD is underestimated)

At this stage, it is being assessed whether PD estimated is significantly higher/lower than the observed default rate (DR). By using a one-sided binomial test it can be effectively assessed whether the final PD value used for RWA computation is not underestimating the risk. Ensuring the p-value is above the chosen significance level indicates that the PD is appropriately estimated, while a p-value below the significance level suggests the need for adjustments to avoid underestimating the risk.

4. CASE STUDY

This section applies the previously outlined conceptual framework across four subsections. The first subsection presents the main concepts necessary for constructing the model. The second subsection details the data collection and processing methodology. The final two subsections focus on the model's construction and evaluation, respectively. The scenarios discussed in Chapter 5 are based on the results presented in this section.

4.1 Developing the model

This section will provide a detailed explanation of the foundational elements and methodologies that substantiate the model's development, ensuring a clear understanding of the principles and assumptions that guide the process.

4.1.1 Scope and Perimeter

The scope refers to the specific portfolio of exposures to which the model applies. In this case, the focus is on Corporates, Large Corporates, and Institutions, excluding Central governments and Central Banks. The decision to choose a Corporate portfolio over a Retail one was made to avoid diving into a larger dataset, as the number of defaults and obligors is typically much higher in Retail portfolios, which could lead to significant issues in data quality controls. However, it is important to note that the dataset used is already complex, dealing with around 100 000 obligors per year. A larger dataset would exacerbate these challenges, further complicating data management and increasing the computational workload, thereby making the calculations significantly more complex and time-consuming. The analysis will cover the period from 2011 to 2022. This choice is constrained by data availability but is realistic as it is following EBA requirements (more than 5 years and contains “good” and “bad” years – low and high defaults years). In subsection 4.2 it is explained with more detail how the data regarding the number of obligors was collected. The Table 1 presents the size of the portfolio per year considered.

Table 1 | Number of obligors per year

Year	Obligors	Year	Obligors
2011	100 000	2017	113 200
2012	97 466	2018	119 208
2013	94 786	2019	123 446
2014	98 743	2020	85 781
2015	101 001	2021	80 127
2016	110 272	2022	102 768

Source: BNPP financial reports and simulated data

For the purpose of this work, this sample is going to be split in three different groups: modelling sample (2011-2018); assessment sample (2011-2019) and the remaining years are going to be used to perform the Backtesting exercises.

4.1.2 Risk Criterion

The choice of the risk criterion is a key part and will impact the whole modelling strategy.

Risk criterion is defined as a binary target which identifies the performing and non-performing obligors. For the granting scores or the behavioural scores, the risk criterion chosen by default is the Basel default. According to EBA, "[…] *a default shall be considered to have occurred […] when either or both of the following have taken place:*

- *The institution considers that the obligor is unlikely to pay its credit obligations to the institution, the parent undertaking or any of its subsidiaries in full, without recourse by the institution to actions such as realising security;*
- *The obligor is more than 90 days past due on any material credit obligation to the institution, the parent undertaking or any of its subsidiaries."*[7]

Using this definition, the Default Rate (DR) comes as the following ratio:

$$DR = \frac{d}{N} \tag{14}$$

where:

- **d** is the number of non-defaulted obligors/facilities at the beginning of the one-year observation period which had a default event during this period;
- **N** is the total number of non-defaulted obligors/facilities at the beginning of the one-year observation period;

In section 4.2, is explained how the DR was collected. Table 2 shows the different default rates observed on the perimeter across the years considered, which are aligned with equation 14.

Table 2 | Default Rate per year

Year	DR	Year	DR
2011	1.03%	2017	1.61%
2012	1.10%	2018	1.43%
2013	1.16%	2019	1.27%
2014	1.04%	2020	1.19%
2015	0.82%	2021	0.90%
2016	1.00%	2022	0.74%

Source: BNPP financial reports

[7] Article 178 | European Banking Authority (europa.eu)

On the development of the model, a cohort (or snapshot) analysis will be used. It corresponds to a picture of all non-defaulted obligors at a snapshot date *t*, for which the occurrences of default are observed between *t* and *t + 12 months*.

4.1.3 Risk Drivers and other variables

By Risk Drivers (RD), one means the explanatory variables allowing the risk differentiation across the sample population with the considered risk parameter. The RD and other variables selected can be of two types:

- Quantitative: Indicate that variables have a numeric format. Quantitative variables can be continuous (amount of the loan) or discrete (number of past unpaid payments).
- Qualitative: Indicate that variables belong to one category of a finite ensemble. Arithmetic operations do not make sense on these variables (level of education, matrimonial status…). They can ben ordinal (risk profile) or nominal (home status).

In the document "ECB guide to internal models"[8], the ECB defines a list of potential risk drivers that should ensure a meaningful differentiation of risk across the obligors:

"[…]it is the ECB's understanding that PD models should perform adequately on economically significant and material sub-ranges of application which are identified […]on the basis of potential drivers for risk differentiation, including the following non-exhaustive list of drivers, where relevant:

(a) […] country, industry, size of obligor, past delinquency (e.g. obligors with delinquency events, i.e. days past due, in the last 12 months), firm age;

(b) […] client type, product type, region, maturity (e.g. original or remaining maturity), type of real estate;

The selection of the right RDs is crucial when constructing the model, but the number of RDs chosen is equally important. It is essential to balance good explanatory power while maintaining an efficient process. To achieve this, 4 RDs were selected based on the previous list.

- Age of the relation between the obligor and the institution (e.g. BNPP). This RD will be referred as Seniority in the following sections.

[8] Article 55, page 75 of ECB guide to internal models (europa.eu)

- Sales.
- Debt Ratio. Ratio between total liabilities over total assets.
- Delinquency. It means that the obligor is past due over 30 days on its financial obligations.

When using the logistic regression, it is common practice to bin the explanatory variables before using them in the model. It serves multiple objectives such as: having a more readable score grid; taking into account non-linear effects of the variables; reduce the impact of extreme values on the estimation of the coefficients, as well as on the prediction based on the score.

Binning must comply with some criterions:

- Classes must be differentiated in risk rate (monotonicity): the binned variable's effect on the risk must be monotonous, meaning that the default rate increases or decreases by category;
- Volumes in each class must be sufficient (minimal volume of at least 50 non-default and default observations, and contain at least 5% of the total population);
- The binning should be in line with the business interpretation of the variable (for delinquency, either an obligor is in failure or not. Thus, this variable will be binned in 2 categories);

Thus, the risk drivers have been binned as follows:

- Seniority (y):
  - $y \in]0,1]$
  - $y \in]1,10]$
  - $y \in]10,15]$
- Turnover (T). The following values are presented in millions of euros:
  - $T \in]0.15,0.5]$
  - $T \in]0.5,1.5]$
  - $T \in]1.5,4]$
  - $T \in]4,7]$
  - $T \in]7,10]$
- Debt Ratio (Total Liabilities over Total Assets):
  - $ratio \in]0,1]$
  - $ratio \in]1,1.5]$
  - $ratio \in]1.5,3]$
- Delinquency.

When choosing the explanatory variables, some measures can be used to have a first look to the discriminatory power of each of them. Among those measures, the following were used: Weight of Evidence (WoE), Informative Value (IV) and Cramer's V.

- **Weight of Evidence**

The WoE[9] is a statistical measure that indicates the predictive power of an independent variable in relation to a dependent variable. Its origins lie in the credit scoring industry, where it is used to assess the distinction between performing and non-performing obligors.

$$\text{WoE} = \ln\left(\frac{\%\text{ of performing obligors}}{\%\text{ non performing obligors}}\right) \tag{15}$$

- **Informative Value**

The IV[9] is a valuable technique for selecting the important variables in a predictive model. It allows for ranking variables according to their relative importance, thereby facilitating an initial understanding of the discriminatory power of the variables in question. Both formulas are presented below.

$$\text{IV} = \sum(\%\text{ of performing obligors} \times \%\text{ of non performing obligors}) \times \text{WoE} \tag{16}$$

- **Cramer's V[9]**

Cramer's V is a measure between 0 (null association) and 1 (perfect association) that indicates how strongly two categorical variables are associated. It corresponds to the normalised version of the Chi-squared.

$$\text{Cramer's V} = \sqrt{\frac{\frac{\chi^2}{n}}{n(k-1)}} \tag{17}$$

where *k* denotes the smaller number between columns or rows. The χ^2 is the chi-square statistic between the risk driver and the default and *n* is the total sample size.

[9] Internal documentation and Weight of Evidence (WOE) and Information Value (IV) Explained (listendata.com)

Besides risk drivers, the RDS is composed of other variables. The following variables were chosen to enhance the readability and organization of the data, even though they are not directly used in the computations or displayed in the results:

- Reference Date (categorical): date of each snapshot (the range of its values is concerned to the following period: 31/12/2011 – 31/12/2022).
- Obligor (numeric): legal entity which is a counterparty to a credit facility (it will be represented as Obligor ID indicating the number associated to each obligor).
- Default Flag (categorical): variable that indicates, for a given snapshot, if the obligor has defaulted over the 12 months considered (it can take the value 0 in case of non-default observed or 1 in case of a default had happened).
- Cluster (categorical): variable represents a risk category within the obligor rating scale of a rating system, where obligors are assigned based on a specific set of rating criteria. Obligors within the same Risk Class cannot be divided into subgroups with significantly different default rates.
- Default Rate (numeric).

4.1.4 Scoring

A key determinant of a PD model's quality is its ability to accurately sort obligors based on their risk to default. As previously discussed, scoring approaches are a method of classifying obligors where each obligor is assigned a score reflecting its risk level. This approach was used for constructing the model. Based on the 4 Risk Drivers selected earlier and in their different categories, it is resulting in 90 possible combinations, meaning 90 or less possible distinct scores.

The process of breaking down categorical variables into distinct categories using dummy variables is a fundamental yet sometimes complex aspect of model construction. Green (2013) recognizes the use of dummies variables as a powerful tool for capturing non-linear relationships between variables and outcomes. This approach allows the model to account for the influence of each category individually, improving the model's ability to differentiate between varying risk profiles.

However, this technique comes with its own set of challenges. Creating dummy variables can significantly increase the dimensionality of a dataset, especially when dealing with variables that have multiple categories. For instance, a variable with four categories would require the creation of three separate dummy variables.

$$I_1 = 1 \text{ if category}_1, 0 \text{ otherwise}$$
$$I_2 = 1 \text{ if category}_2, 0 \text{ otherwise}$$
$$I_3 = 1 \text{ if category}_3, 0 \text{ otherwise}$$

In this setup, one dummy is assigned to each category except the fourth, which is identified by the absence of all three dummies, known as the reference category. This "expansion" can lead to some problems like overfitting or multicollinearity, where the newly created variables become highly correlated with one another. This process is then applied to the remaining risk drivers.

A critical decision in this process is selecting the reference category. This category serves as the baseline against which all other categories are compared, offering an easier way of interpreting the results of the model. To run the logistic regression, the less risky category of each risk driver was chosen as the reference. When running the regression, to exclude the reference category prevents the "dummy variable trap," a situation where including a dummy variable for every RD would cause perfect multicollinearity, causing incorrect calculations of regressions coefficients and their corresponding p-values.

After performing the regression, the coefficients obtained must be analysed across several aspects:

- Significance of the coefficients (Wald tests):[10] Wald tests are applied to the model's coefficients for each category transformed into dummies. If the p-value exceeds a 5% threshold, the coefficient is considered non-significant (indicating that the corresponding dummy has no effect). In such cases, it may be necessary to re-bin the variable or group non-significant attributes with those having similar coefficients or risk rates, provided the business meaning of the variable is maintained.
- Coherence between the coefficients and risk rates: The coefficients assigned to the categories of a variable should align with the observed risk rates (coefficients and risk rates should follow the same trend: either increasing or decreasing). If there is an inversion, it could indicate redundancy among variables within the model, necessitating further investigation. Potential solutions include regrouping categories, redefining variable binning, or cross-referencing variables.

[10] Based on BNPP internal documentation.

- Coherence with the business meaning of the variables: The coefficients should be consistent with the business interpretation of the variables, ensuring the model's practical relevance.
- Correlation between the variables in the model: It is essential to check for correlations among the variables using Cramer's V, to avoid multicollinearity.
- Volume constraints: The volume constraints for each category within the variables must be adhered to throughout the model. These constraints require that each category includes at least 50 non-default and default observations and represents at least 5% of the total population.

After addressing the previous steps, the next phase involves normalizing and centring the coefficients. The raw coefficient values from the regression can be challenging to interpret, but this can be addressed using a straightforward technique: coefficient normalization. This technique applies a linear transformation to the raw scores, making them easier to interpret.

Centring is used to ensure that normalized scores are balanced and comparable across different variables[10]. In the context of a scoring model, centring adjusts the scores so that the intercept is effectively distributed among the variables, eliminating the need for an intercept while maintaining the model's interpretability. By integrating the intercept's effect into the coefficients of other variables, the scoring model becomes more balanced and easier to understand.

Although confidentiality restrictions prevent the use of the original formula for normalizing the scores or calculating the adjustment value, similar formulas are applied to achieve the same objective.

4.1.5 Clusters

After obtaining the score, the next step is to build Risk Classes, utilizing the K-means algorithm as mentioned. These Risk Classes are defined by the ECB as *"a subset of obligors to which the same PD is applied for the calculation of regulatory capital requirements*.".[11]

Ikotun et al. (2022), stated that recent advancements in scientific data collection techniques in the big data era have enabled the systematic gathering of vast amounts of

[11] *ECB Guides to Internal Models, February 2024*. From ECB guide to internal models (europa.eu), article 100.

data from various collection sites. Alongside this, there has been significant growth in the development of data analysis methods, with the K-means algorithm remaining one of the most popular and straightforward clustering techniques. However, the K-means algorithm faces several challenges that can impact its clustering performance. One major issue is the need for the user to specify the number of clusters in the dataset beforehand, with initial cluster centres selected randomly. Additionally, the algorithm's performance is highly sensitive to the initial cluster selection, and determining the optimal number of clusters for large datasets can be complex and challenging.

Using the score obtained after scoring, the first task is to classify each possible combination of risk drivers based on this score. There are 90 different possible scores, which are ranked from the least risky to the riskiest.

To initialize the K-means algorithm, the combinations are randomly assigned to clusters or classes, which is a crucial step for minimizing the number of iterations. Based on these initial clusters, the centroids are set, typically calculated as the average score of each class, which is the most common method. Next, the Euclidean distance between each combination and each centroid is calculated. If the distance is minimized for the centroid of the cluster to which the combination is already assigned, the combination remains in that cluster. If not, it is reassigned to the cluster where the distance to the centroid is minimized, either moving to a higher or lower RC. The algorithm reaches convergence when two consecutive iterations result in no changes to the composition of clusters.

The final step in the model-construction phase consist in ensure that clusters are heterogeneous. Thus, it is necessary to conduct the Welch Test described before.

4.2 Data collection and treatment

Data on the number of obligors was sourced from the public annual financial reports of BNPP, specifically following the Basel III Pillar III disclosures. However, the level of detail in these reports has varied over time, meaning that precise data on the number of obligors is only available for the years 2014 to 2022. To address this gap, the number of obligors for the years 2012 and 2013 was randomly generated based on the standard deviation observed in the 2014-2022 data. For 2011, it was decided to fix the number of obligors at 100 000 as initial value since it is a value inside the range of obligors verified in the available reports.

The DR was also collected from the public annual financial report of BNPP and as for the number of obligors, some treatments had to be performed to get the DR across the years considered.

- Between 2017 and 2022, the reports discretize in a deeper way the composition of the portfolio (both Retail and Corporate) stating the number of obligors assigned to each level of DR considered. To calculate the DR, one simply needs to sum the number of obligors that defaulted and divide by the total number of obligors in the considered portfolio.
- For the period between 2014 and 2016, the disclosed information did not specify the number of defaulted obligors. However, data on the total number of obligors in the portfolio and their DR was available in BNPP public annual financial reports. Since the DR for Institutions, Corporates, and Large Corporates was presented separately, the overall DR was obtained by calculating the weighted average of the DRs for each category for the three years.
- For the period between 2011 and 2013, the same process used for 2014-2016 was applied, but with randomly generated numbers of obligors.

Ensuring the reliability of the data within the RDS during both data collection and simulation is of great importance. Therefore, various assessments must be conducted on the variables across the following dimensions:

- Timeliness: The data must be current, meaning the values should be up to date. This is a check that only concerns the DR and the obligors since these are the only variables which are being extracted. As both variables derived from publicly available financial reports, they are inherently current.

- Uniqueness: The data should be free from duplication, particularly when aggregating or applying filters to the source data. This verification is focused on primary keys. While constructing the RDS, no obligor was recorded more than once in any given year.

- Completeness: The RDS must contain all necessary values for the required attributes. Given that the data is directly sourced from financial reports, and the remaining data is simulated accordingly (next section), there are no missing values in the RDS.

- Consistency: The data must be appropriately formatted within the relevant attributes. For example, 'Reference Date' should be in the correct date format.

In a standard process, it is also important to account for potential issues such as outliers or trends in the data. However, due to the controlled construction of the RDS, tests or treatments for these issues are unnecessary, as such situations are prevented from occurring.

4.3 Building the model

This subsection will be subdivided into two parts. The first one it will expose the construction process of Risk Drivers. In the second, it will be demonstrated all the process behind the score grid of the obligors as well their allocation to the clusters.

4.3.1 Risk drivers

As discussed in subsection 4.1.3, binning variables in categories is a common practice when using logistic regression. This process must follow established guidelines to ensure that the categories effectively differentiate risk and align with the economic interpretation of the variable in question.

Due to confidentiality and client data protection concerns, true values could not be used when populating the risk drivers or distributing them among the different categories. The following strategy was adopted:

A variable called "composition" was defined, and the next step involved assigning a DR to each of the different risk driver categories. By multiplying these two variables, we obtain the percentage of obligors within each category who have defaulted. The sum of these percentages for each risk driver must equal the DR for the corresponding year. The values assigned to these variables are determined through a combination of trial and error and expert judgment.

For example, with the first risk driver, it is reasonable to expect that, *ceteris paribus*, obligors with a more recent relationship with the institution will have a higher DR than those with a longer-standing relationship. Simultaneously, it makes sense to assume that most clients have established some level of longevity with the institution, meaning the majority are not new or recent clients. Conversely, with the last risk driver, the expectation is reversed. An institution would likely have a smaller proportion of

customers in arrears, so most obligors are expected to have a Flag_Delinquency = 0, while those with Flag_Delinquency = 1 are more likely to default.

Considering these factors, obligors were randomly distributed among the different categories based on their composition and DR variables. Below is an example illustrating this process.

Table 3 | General information of 2017

Year	DR	Number of obligors	Number of defaults
2017	1.61%	113 200	1 823

Source: Simulated data

Table 4 | Computation of defaults on seniority for the year of 2017

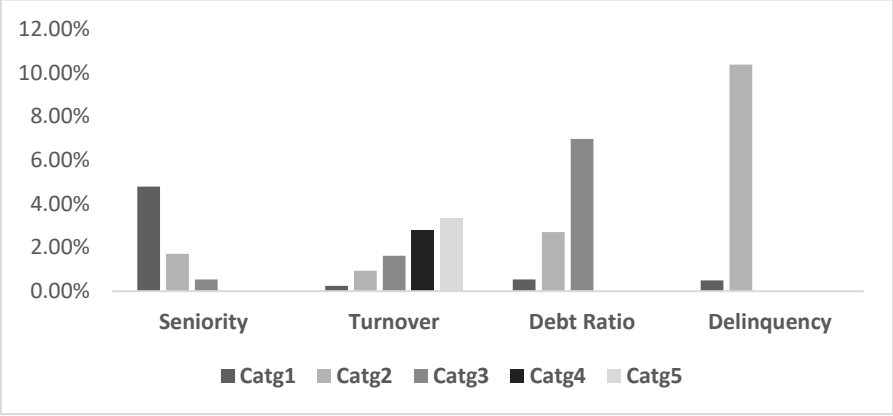
Composition	Category	DR	Distribution of defaults	Expected number of defaults
10%	]0,1[	4.88%	$0.1 \times 0.0488 = 0.49\%$	$0.49\% \times 113200 = 552$
55%	[1,10[	1.71%	$0.55 \times 0.0171 = 0.94\%$	$0.94\% \times 113200 = 1065$
35%	[10,15]	0.52%	$0.35 \times 0.0052 = 0.18\%$	$0.18\% \times 113200 = 206$
Total 100%	-	-	1.61%	1 823

Source: Simulated data

The logic behind the construction of Table 4 can be applied to calculate the number of obligors who did not default for a considered year. Instead of using the DR it is used (1 - DR) and the remaining process is equivalent.

It is equally important to confirm that the binning properties are verified, which will be examined in the following graph.

Graph 1 | Monotonicity of risk drivers for the year 2017



Source: Simulated data

As shown in Graph 1, the categories effectively differentiate risk across each explanatory variable. Riskier categories exhibit a higher default rate, while less risky

categories show a lower default rate. Additionally, the absence of inversion points indicates the presence of monotonicity in the default rates across the categories. The default rates also align with the expected business logic; for example, an obligor's probability of default increases proportionally with their debt ratio, assuming all the other factors remain constant.

The completion of the risk drivers' construction depends on analysing the IV, WoE, and Cramer's V values. The following results were obtained using equations 15-17.

Table 5 | WoE and IV of risk driver seniority

Category	WoE	IV
]0,1[	3.13	0.29
[1,10[	4.39	2.36
[10,15]	6.21	2.16
-	-	$\sum$ IV = 4.81

Source: Simulated data

The preceding table displays the results for these indicators for Seniority. Both IV and WoE show significant predictive power, with notably high values, except for the initial category. Consequently, it is recommended to retain this binning approach and the variable in the model. Similar results were obtained for the remaining risk drivers, leading to analogous conclusions (see Tables 23-25 in Appendix).

Table 6 | Cramer's V of RDs

Risk Driver	Cramer's V
Seniority	0,07
Turnover	0,96
Debt Ratio	0,16
Delinquency	0,24

Source: Simulated data

As shown in Table 6, the risk driver Seniority has a low Cramer's V value of 0.07, which is not consistent with the WoE and IV values. In contrast, the Risk Driver Turnover shows a strong correlation with the dependent variable, with a coefficient of 0.96, reflecting particularly strong explanatory power. The other two Risk Drivers, Debt Ratio and Delinquency, have correlation coefficients of 0.16 and 0.24, respectively, indicating a reasonable degree of explanatory power

Additionally, Cramer's V values, which measure the association between risk drivers, were calculated to preliminarily assess the potential for multicollinearity. As seen in Table 26 in Appendix, all Cramer's V values are below 0.027, except for the Turnover and Debt

Ratio correlation, which is 0.278. This suggests that the variables are not highly correlated.

4.3.2 Cluster Construction

As previously mentioned, the first step in assigning obligors to clusters involves assigning them a score that reflects the level of risk they pose to the bank. A logistic regression analysis is conducted, according to equations 1-10, with the results presented below.

Table 7 | Score grid

Variable	Category	Coefficient	P-value	Number of obligors	Percentage of obligors	Number of defaults	Risk Rate
Intercept	-	-12.2643	< 0.01	-	-	-	-
Seniority	]0,1[	2.8754	-	83 197	9.97%	3 471	4.17%
	[1,10[	1.2386	< 0.01	459 059	55.02%	5 632	1.23%
	[10,15]	0	< 0.01	292 150	35.01%	585	0.20%
Turnover	]0.15,0.5[	0	< 0.01	50 022	6.02%	69	0.14%
	[0.5,1.5[	0.4260	< 0.01	250 569	30.0%	1 136	0.45%
	[1.5,4.0[	0.4805	< 0.01	292 064	35.0%	3 459	1.18%
	[4.0,7.0[	0.9244	< 0.01	174 965	20.97%	3 134	1.79%
	[7.0,10.0[	1.1871	-	66 786	8.01%	1 890	2.83%
Debt Ratio	[0.0,1.0[	0	< 0.01	583 945	69.98%	1 152	0.20%
	[1.0,1.5[	1.1367	< 0.01	167 251	20.04%	3 213	1.92%
	[1.5,3.0[	2.7259	-	83 210	9.98%	4 923	5.92%
Delinquency	0	0	< 0.01	750 983	90.0%	2 356	0.31%
	1	3.3930	-	83 423	10.0%	7 332	8.79%

Source: Simulated data

Based on the criteria outlined in the previous section, the following conclusions can be drawn:

All categories across the various risk drivers exhibit significant coefficients. For risk drivers where the risk rate increases across categories, the coefficients decrease correspondingly, maintaining consistency with their attributes. No changes were made to the thresholds during the regression analysis, ensuring that the variables and their categories retain their original interpretation. Additionally, the analysis adheres to the requirement of the regulation that each category must include at least 50 non-default and default observations and represent a minimum of 5% of the total population. These constraints are met, as evidenced by the results.

The next step involves normalizing and centring the regression coefficients to derive the final score for each category. Due to confidentiality, the internal formulas used for these processes cannot be disclosed. However, the following alternative formulas will be employed[10]:

- **Normalization**:

$$\text{Reference value} + \text{regression coefficient} \cdot \frac{20}{\ln(2)} \quad (17)$$

- **Centring:**

$$\text{Normalized coefficient} + \frac{\text{Normalized intercept coefficient}}{2} \quad (18)$$

Table 8 | Normalized and Centred score grid

Variable	Coefficient	Category	Normalised coefficients	Centring
Intercept	-12.2643	-	46	0
Seniority with the institution	2.8754	]0,1[	83	106
	1.2386	[1,10[	36	59
	0	[10,15]	0	23
Turnover	0	]0.15,0.5[	0	23
	0.4260	[0.5,1.5[	12	35
	0.4805	[1.5,4.0[	14	37
	0.9244	[4.0,7.0[	27	50
	1.1871	[7.0,10.0[	34	57
Debt Ratio	0	[0.0,1.0[	0	23
	1.1367	[1.0,1.5[	33	56
	2.7259	[1.5,3.0[	79	102
Delinquency	0	0	0	23
	3.3930	1	98	121

Source: Simulated data

After normalization and centring by applying equations 17-18, each obligor is given a score by combining the scores for each category. An obligor with less than one year of seniority, a turnover of less than 500 000 euros, a debt ratio of 1.3, has not failed to fulfil their obligations within the past 30 days receive a final score of 208 (106+23+56+23).

There are 90 distinct possible combinations, which must be ranked in order of risk, from the least risky to the riskiest. Once ranked, these combinations are used to initialize the K-means clustering algorithm (previous section). The process was initialized with 9 clusters, and 5 iterations were performed to verify the convergence criteria. The most significant changes occurred between clusters 2 and 3, where the size of cluster 2 decreased as some categories were reassigned to cluster 3. A similar adjustment occurred between clusters 5 and 6. Below is a summary of the final output.

Table 9 | Final outcome of K-mean cluster algorithm

Cluster	Score	Defaults	Obligors	DR
1	92-142	167	437 543	0.0382%
2	152-175	351	165 525	0.2121%
3	183-209	525	124 846	0.4205%
4	217-226	465	25 189	1.8460%
5	234-254	1 160	48 860	2.3741%
6	257-273	1 074	10 697	10.0402%
7	281-307	1 543	14 178	10.8831%
8	317-340	3 063	5 850	52.3590%
9	352-386	1 340	1 718	77.9978%

Source: Simulated data

The final step in the model-building process is to ensure that the clusters accurately differentiate risk, meaning they must be heterogeneous. To verify this, the Welch test was applied and the results confirm the heterogeneity of all clusters.

4.4 MODEL ASSESSMENT

To meet the standards set by the ECB for internal models, Risk Class must demonstrate proficiency in several key areas before being validated: Stability, Discriminatory Power, Heterogeneity, and Accuracy. This evaluation process, known as a validation assessment, aims to ensure that the model accurately reflects these dimensions. To test the model, a new sample covering the entire modelling period, along with an additional year (2019), is required.

Table 10 | Assessment scenario: dimension's outcomes

Dimension	Test	Modelling	Assessment
Stability	PSI	1.91%	2.05%
Discrimination	AR	0.89	0.88
Heterogeneity	Welch test	No lack of heterogeneity	[4,5], [7,8]
Calibration	2-sided Binomial Test	-	Well calibrated
Conservatism	1-sided Binomial Test	-	Globally conservative

Source: Simulated data

Model stability is measured by the Population Stability Index (PSI), as presented on equation 12. As seen in Table 10, the PSI value is 2.05%, which, according to the defined threshold, indicates that the distribution of obligors across the clusters in the validation sample closely matches the distribution used in the modelling sample.

Discriminatory Power allow to verify if the ranking of obligors effectively distinguishes between those with higher and those with lower risk profiles. The model's discriminatory capacity is evaluated using the AR metric, shown in equation 13, which

shows a value of approximately 0.88, therefore aligned to the one observed during the modelling phase.

Heterogeneity is assessed using the Welch Test as posted in equation 10. By performing it, it is ensured that all the clusters are heterogeneous, meeting the criteria.

Clusters must also ensure that its estimates align closely with observed data, a process known as calibration. A binomial two-sided test was performed as shown in Table 11. For clusters 4, 6, 8 and 9 the PD is not well calibrated however, at model level, the null hypothesis is rejected which means that the estimated PD is in an acceptable range.

Table 11 | Model's calibration

Cluster	Obligors	PD	PD (MoC)	DR	p-value	Assessment
1	470 786	0.05%	0.05%	0.04%	0.482	Not reject H_0
2	178 253	0.26%	0.27%	0.24%	0.255	Not reject H_0
3	133 751	0.49%	0.52%	0.46%	0.144	Not reject H_0
4	28 617	4.20%	4.29%	1.98%	0.000	Reject H_0
5	52 649	2.69%	3.19%	2.59%	0.194	Not reject H_0
6	11 943	11.61%	11.99%	10.47%	0.000	Reject H_0
7	15 329	11.99%	13.00%	11.70%	0.268	Not reject H_0
8	6 685	67.06%	72.09%	52.55%	0.000	Reject H_0
9	1 928	80.96%	89.54%	77.85%	0.001	Reject H_0
Total	899 941	1.47%	1.79%	1.25%	0.476	Not reject H_0

Source: Simulated data

For the purposes of this work, it is relevant to provide an example of MoC. However, calculating it requires resources and tools that are only accessible to institutions regulated by the ECB. Table 12 displays the MoC values, which are included for illustrative purposes in this context and are in line with the ones used during my internship.

Table 12 | MoC values

Cluster	MoC	PD	PD (including MoC)
1	2,9%	0.05%	0.05%
2	4,7%	0.26%	0.27%
3	5,2%	0.49%	0.52%
4	2,1%	4.20%	4.29%
5	18,70%	2.69%	3.19%
6	3,2%	11.61%	11.99%
7	8,4%	11.99%	13.00%
8	7,5%	67.06%	72.09%
9	10,6%	80.96%	89.54%
Total	-	1.28%	1.54%

Source: Simulated data

This leads to the following final PD estimate for the model: 1.54% which is higher than the LRA DR (1.28%). The model underestimated the risk.

5. BACKTESTING EXERCISES

As previously discussed, Backtesting tests rely on four key dimensions: Discrimination, Stability, Calibration accuracy and Conservatism.

This subsection will present 3 scenarios as examples of Backtesting exercises designed to assess the model's fit across different dimensions. A scenario, in this context, refers to the simulation of new observed data based on pre-defined hypothesis. In each scenario, 3 years of observations will be simulated. Indeed, the problem in generating only one additional observation year is that the TTC results would not be affected. The first scenario is "neutral", the idea is to generate new observations by following the same parameters than the ones used earlier to generate the modelling RDS. In the second scenario, a more adverse environment is simulated, with an higher DR than the one which was effectively observed, in order to evaluate the model's robustness under riskier conditions. Finally, in the third scenario, a less risky environment is simulated.

5.1 Scenario 1

In this scenario, the risk of the portfolio will remain consistent with previous analyses as well as the number of obligors and their distribution across the clusters. A Backtesting exercise will be performed for the year 2022. The objective is to observe the reaction of the model to this new, but similar, data through the results of the backtesting tests.

Table 13 | Scenario 1: Dimension's assessment outcomes

Dimension	Test	Modelling	2022
Stability	PSI	1.91%	1.58%
Discrimination	AR	0.89	0.91
Heterogeneity	Welch test	No lack of heterogeneity	[4,5], [7,8]
Calibration	2-sided Binomial Test	-	Well calibrated
Conservatism	1-sided Binomial Test	-	Globally conservative

Source: Simulated data

To assess the model's Stability and Discriminatory Power, PSI, AR, IV, WoE, and Cramer´s V indicators are used, according to equations 12, 13 and 15-17, respectively. Given the heavy computation workload when computing the IV, WoE and Cramer's V, this will be the only scenario where these values are going to be displayed.

Table 14.1 | Scenario1: Cramer's V between RDs

Year	1_2	1_3	1_4	2_3	2_4	3_4
2022	0.005	0.01	0.012	0.008	0.015	0.028

Table 14.2 | Cramer's V between RDs and the target variable

Year	1_t	2_t	3_t	4_t
2022	0.081	0.01	0.151	0.190

Source: Simulated data

As seen in Table 14.1, the low Cramer´s V values between the risk drivers indicate that they are not highly correlated to each other's, ensuring that the model does not encounter issues related to multicollinearity. In Table 14.2, it is evident that the risk drivers show a small correlation with the target variable, consistent with the values obtained during the modelling period.

The PSI value - 1.58% - indicates that the obligor's distribution across the clusters is consistent with the modelling phase. This conclusion is further supported when comparing the percentage of obligors per cluster between the modelling period and 2022, showing no significant discrepancies.

Table 15 | Scenario 1: Obligor's distribution: Modelling (M) vs 2022

Cluster	% obligors (M)	% obligors in 2022
1	52.44%	51.76%
2	19.84%	19.36%
3	14.96%	13.41%
4	3.02%	5.15%
5	5.86%	5.72%
6	1.28%	1.78%
7	1.70%	1.66%
8	0.70%	0.95%
9	0.21%	0.20%

Source: Simulated data

As observed, the AR (TTC) is 0.91. This represents an increase of 2.1% compared to the modelling period. This value remains close to the levels observed during the modelling phase, indicating that the model retains strong predictive power according to the defined threshold.

Next, it is important to assess whether the clusters remain heterogeneous.

In terms of heterogeneity, it is observed that in 2022, there are two pairs of clusters with lack of heterogeneity [4,5] and [7,8]. A potential solution to this issue is to adjust the clustering, allowing certain categories to move from one cluster to another.

Calibration accuracy and Conservatism are the final dimensions to assess in a Backtesting exercise.

Table 16 | Scenario 1: Calibration (CA) and conservatism (CO)

Cluster	Obligors	PD	PD (MoC)	DR	p-value (CA)	Assessment (CA)	p-value (CO)	Assessment (CO)
1	609 339	0.05%	0.05%	0.04%	0.482	Not reject H_0	0.998	Not reject H_0
2	230 400	0.26%	0.27%	0.23%	0.409	Not reject H_0	0.995	Not reject H_0
3	170 051	0.49%	0.52%	0.45%	0.416	Not reject H_0	1.000	Not reject H_0
4	42 610	4.20%	4.29%	1.77%	1.131	Not reject H_0	1.000	Not reject H_0
5	68 063	2.69%	3.19%	2.47%	0.389	Not reject H_0	1.000	Not reject H_0
6	16 722	11.61%	11.99%	9.15%	0.515	Not reject H_0	1.000	Not reject H_0
7	19 583	11.99%	13.00%	10.95%	0.348	Not reject H_0	1.000	Not reject H_0
8	9 204	67.06%	72.09%	46.85%	0.228	Not reject H_0	1.000	Not reject H_0
9	2 478	80.96%	89.54%	72.64%	0.000	Reject H_0	1.000	Not reject H_0
Total	1 168 450	1.52%	1.64%	1.18%	0.579	Not reject H_0	1.000	Not reject H_0

Source: Simulated data

Concerning Calibration, it is possible to notice that the model is globally well calibrated. When looking at cluster level, the null hypothesis is not rejected for all the clusters, except for cluster 9.

The model is globally conservative (the final PD is higher than the LRA DR – 1.64% vs 1.18%). The H_0 is not rejected for all the clusters which means that for all the clusters the final PD of each is overestimating the risk, and therefore are conservative.

To sum up with this first scenario, the model kept its stability and discriminatory power when compared with the modelling phase. Regarding heterogeneity, two pairs of clusters present similar levels of risk. The PD estimated is well calibrated. The model is conservative, clearly overestimating the risk.

5.2 Scenario 2

In this scenario, the portfolio's risk will increase compared to the previous analyses. As in Scenario 1, this increase in risk will only impact the DR values and not in the number and distribution of the obligors.

Table 17 | Scenario 2: Dimension's assessment outcomes

Dimension	Test	Modelling	2022
Stability	PSI	1.91%	1.63%
Discrimination	AR	0.89	0.68
Heterogeneity	Welch test	No lack of heterogeneity	No lack of heterogeneity
Calibration	2-sided Binomial Test	-	Well calibrated
Conservatism	1-sided Binomial Test	-	Globally conservative

Source: Simulated data

The TTC PSI value for 2022 – 1.63% - stills indicates that the obligor's distribution across the clusters is consistent with the modelling phase. This conclusion is further

supported by comparing the percentage of obligors per cluster for each year compared to the modelling period, which shows no significant differences (except in cluster 4) as shown in the table below.

Table 18 | Scenario 2: Obligor's distribution: Modelling (M) vs 2022

Cluster	% obligors (M)	% obligors in 2022
1	52.44%	51.71%
2	19.84%	19.32%
3	14.96%	13.43%
4	3.02%	5.25%
5	5.86%	5.83%
6	1.28%	1.77%
7	1.70%	1.62%
8	0.70%	0.89%
9	0.21%	0.18%

Source: Simulated data

The TTC AR value for 2022 is 0.68. Despite a 24% decrease compared to the modelling period, the value still demonstrates that the model effectively differentiates obligors based on their risk profiles. Comparing this value with the one obtained in the previous scenario, it can be observed that the model can adjust its predict power more easily in less risk environments rather than in riskier environments.

Regarding heterogeneity, contrary to what was observed in the previous scenario, no lack heterogeneity in pair of clusters was found, which means that the model stills correctly differentiate the risk across the clusters.

Table 19 | Scenario 2: Calibration (CA) and Conservatism (CO) TTC

Cluster	Obligors	PD	PD (MoC)	DR	p-value (CA)	Assessment (CA)	p-value (C)	Assessment (C)
1	609 702	0.05%	0.05%	0.10%	0.000	Reject H_0	0.00	Reject H_0
2	229 992	0.26%	0.27%	0.43%	0.000	Reject H_0	0.00	Reject H_0
3	169 950	0.49%	0.52%	0.78%	0.000	Reject H_0	0.00	Reject H_0
4	42 715	4.20%	4.29%	2.65%	0.000	Reject H_0	1.00	Not reject H_0
5	68 263	2.69%	3.19%	3.16%	0.000	Not reject H_0	0.681	Not reject H_0
6	16 753	11.61%	11.99%	10.30%	0.000	Not reject H_0	1.00	Not reject H_0
7	19 680	11.99%	13.00%	12.00%	0.972	Not reject H_0	0.998	Not reject H_0
8	9 101	67.06%	72.09%	46.94%	0.000	Reject H_0	1.00	Not reject H_0
9	2 460	80.96%	89.54%	73.86%	0.000	Reject H_0	1.00	Not reject H_0
Total	1 168 616	1.52%	1.64%	1.40%	0.390	Not reject H_0	1.00	Not reject H_0

Source: Simulated data

Regarding calibration, the model is well calibrated at model level but when looking at cluster level, only for clusters 5, 6 and 7, the PD is well calibrated. It can be noticed that 61.5% of the obligors belong to clusters that are underestimating the risk.

The model is globally conservative (the final PD is higher than the LRA DR – 1.64% vs 1.40%). However, 86.4% are in clusters that underestimate the risk. Clusters 1, 2 and 3 fail to present a final PD higher than the observed DR.

As a conclusion, the model kept its stability and capability of well differentiate the obligors according to their risk profile, but the value of the AR decreased compared to the value obtained during the modelling phase. As in the modelling phase, no lack of heterogeneity was detected among the clusters. The addition of a higher risk environment did not see to affect the global results regarding calibration and conservatism.

5.3 Scenario 3

In this scenario, the portfolio's risk will decrease compared to previous analyses. As for the previous scenarios, the number of obligors and their repartition in the clusters will remain stable. Thus, this reduction in risk will only impact the DR values.

Table 20 | Scenario 3: Dimension's assessment outcomes

Dimension	Test	Modelling	2022
Stability	PSI	1.91%	16.59%
Discrimination	AR	0.89	0.94
Heterogeneity	Welch test	No lack of heterogeneity	[4,5]
Calibration	2-sided Binomial Test	-	Globally not well calibrated
Conservatism	1-sided Binomial Test	-	Globally conservative

Source: Simulated data

The TTC PSI value for 2022 – 16.59% - is clearly not aligned with the value obtained in the modelling phase. This can be explained by the lack of obligors on cluster 4, which by itself has a PSI of 14.57%.

Table 21 | Scenario 3: Obligor's distribution: Modelling vs 2022

Cluster	% obligors (M)	% obligors in 2022
1	52.44%	55.35%
2	19.84%	22.83%
3	14.96%	12.40%
4	3.02%	0.02%
5	5.86%	6.96%
6	1.28%	0.66%
7	1.70%	1.24%
8	0.70%	0.36%
9	0.21%	0.18%

Source: Simulated data

As observed in Table 16, the AR value is 0.94. This represents an increase of 5.61% comparing to the modelling period. Thus, the model kept its good discriminatory power.

Regarding heterogeneity, lack of heterogeneity was identified in the [4;5] pair of clusters which may be explained by the lack of observation in cluster 4, given the decrease of obligors verified. As previously suggested, a possible solution is to adjust the grading, allowing certain categories to move from one cluster to another.

Table 22 | Scenario 3: Calibration (CA) and Conservatism (CO) TTC

Cluster	Obligors	PD	PD (MoC)	DR	p-value (CA)	Assessment (CA)	p-value (C)	Assessment (C)
1	627 147	0.05%	0.05%	0.04%	0.000	Reject H_0	1.000	Not reject H_0
2	237 553	0.26%	0.27%	0.24%	0.048	Reject H_0	0.999	Not reject H_0
3	160 049	0.49%	0.52%	0.46%	0.117	Not reject H_0	0.999	Not reject H_0
4	29 408	4.20%	4.29%	2.12%	0.000	Reject H_0	1.000	Not reject H_0
5	65 470	2.69%	3.19%	2.26%	0.000	Reject H_0	1.000	Not reject H_0
6	13 163	11.61%	11.99%	11.07%	0.051	Not reject H_0	0.999	Not reject H_0
7	20 580	11.99%	13.00%	10.90%	0.000	Reject H_0	1.000	Not reject H_0
8	12 020	67.06%	72.09%	36.02%	0.000	Reject H_0	1.000	Not reject H_0
9	3 226	80.96%	89.54%	60.57%	0.000	Reject H_0	1.000	Not reject H_0
Total	1 168 616	1.52%	1.64%	1.17%	0.000	Reject H_0	1.000	Not reject H_0

Source: Simulated data

Regarding calibration, the results are not satisfactory. At model level, the PD estimated is not well calibrated and only clusters 3 and 6 (representing 14.82% of the population) present satisfactory results concerning calibration.

In relation to Conservatism, the results are also aligned with the first scenario. The model is globally conservative (the final PD is higher than the LRA DR – 1.79% vs 1.18%). The H_0 is not rejected for all the clusters which means that all PD estimated are overestimating the risk, and therefore are conservative. The results for Conservatism were already expected since this is a less risk scenario than the first one where this final assessment could already be watched.

To conclude, the model did not keep its stability, even though that this lack of stability is explain by the big decrease regarding the number of obligors in cluster 4. Regarding the discriminatory power, the model kept its ability to correctly differentiate the riskier obligor from the less risky. A lack of heterogeneity was detected in the pair o clusters [4,5]. The reduction of risk lead to some problems regarding calibration. The PD estimated failed to be well calibrated in 7 of the 9 clusters and it is also not well calibrated at model level. Although, regrading conservatism, the results are similar when comparing with scenario 1. The model is conservative, clearly overestimating the risk.

6. Conclusion

This report has focused on the Internal Ratings-Based (IRB) models, emphasizing the risk parameter Probability of Default (PD) and its crucial role in the calculation of Risk-Weighted Assets (RWA). The aim was to construct a PD Model, validate it through Backtesting exercises, and ensure its alignment with regulatory requirements.

The development of the PD Model followed a structured approach, using logistic regression combined with K-means clustering to create distinct Risk Classes, each with a specific assigned PD. This process included the development of a scoring system to rank obligors by risk. The integration of the Margin of Conservatism (MoC) ensured a buffer against potential underestimations of risk, enhancing the model's reliability.

The Backtesting framework was presented, focusing on four key dimensions: Stability, Discriminatory Power, Calibration accuracy and Conservatism.

Three distinct Backtesting scenarios were simulated to test the model's robustness under varying conditions. Scenario 1 represented a neutral environment, where the portfolio's risk and distribution remained consistent with modelling data. Scenario 2 simulated an adverse environment with higher Default Rates (DR), testing the model's resilience under riskier conditions. Scenario 3 presented a less risky environment with lower DRs, assessing the model's performance under these circumstances.

In Scenario 1, the model kept stability and discriminatory power, but two pairs of clusters showed similar risk levels, and each cluster had calibration issue. The model was conservative, generally overestimating risk. In Scenario 2, the model remained stable and differentiated obligors well, despite a significant decrease in AR. No heterogeneity issues were detected, but three clusters presented lack of conservatism. In Scenario 3, the model preserved its stability and discriminatory power, with a slight improvement in AR. Some shifts in obligors' distribution resulted in a slightly higher final PD value.

My internship provided me hands on experience with real world applications of the theoretical concepts I have studied, particularly in the field of statistics. This practical exposure has enhanced my analytical skills and ability to solve problems. The insights gained during this period gave the confidence to keep exploring the financial world.

In conclusion, this report successfully developed and validated a PD model that aligns with regulatory standards and demonstrates robustness across different risk environments.

REFERENCES

1. Bandyopadhyay A. Approaches for Measuring Probability of Default (PD). In: *Managing Portfolio Credit Risk in Banks*. Cambridge University Press; 2016:111-136.
2. Bank for International Settlements (2014). *History of the Basel Committee*. [online] bis.org. Available at: https://www.bis.org/bcbs/history.htm.
3. Bhalla, D. (n.d.). *Population Stability Index and Characteristic Analysis*. ListenData, https://www.listendata.com/2015/05/population-stability-index.html [Acessed 10 August 2024].
4. Castermans, G., et al. "An Overview and Framework for PD Backtesting and Benchmarking." *The Journal of the Operational Research Society*, vol. 61, no. 3, 2010, pp. 359–73. *JSTOR*, http://www.jstor.org/stable/40540263.
5. ECB guide to internal models. (2019). Available at: https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.guidetointernalmodels_consolidated_201910~97fd49fb08.en.pdf.
6. Europa.eu. (2017). *Guidelines on PD estimation, LGD estimation and treatment of defaulted assets | European Banking Authority*. [online] Available at: https://www.eba.europa.eu/activities/single-rulebook/regulatory-activities/model-validation/guidelines-pd-estimation-lgd.
7. Greene, W.H. (2003). *Econometric Analysis*.
8. Ikotun, A.M., Ezugwu, A.E., Abualigah, L., Abuhaija, B. and Heming, J. (2022). K-means Clustering Algorithms: a Comprehensive Review, Variants Analysis, and Advances in the Era of Big Data. *Information Sciences*, 622(622). doi:https://doi.org/10.1016/j.ins.2022.11.139.
9. Martin, K. and Program, A. (n.d.). *Dummy Coding for Dummies*. [online] Available at: https://www.lexjansen.com/wuss/2010/analy/3005_2_ANL-Martin.pdf [Accessed 7 July 2024].
10. Open Risk Academy (2024). Open Risk Manual. Open Risk Manual.

11. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: Logistic regression. Perspect Clin Res. 2017 Jul-Sep;8(3):148-151. doi: 10.4103/picr.PICR_87_17. PMID: 28828311; PMCID: PMC5543767.

12. Renou V., PD Mid Corporate: Model Documentation, RISK FRB, 2020. BNPP internal documentation.

13. Soch, Joram, et al. (2024). StatProofBook/StatProofBook.github.io: The Book of Statistical Proofs (Version 2023). https://doi.org/10.5281/ZENODO.4305949

14. The bank for a changing world. [s.l: s.n.]. Available in: <https://invest.bnpparibas/en/document/universal-registration-document-2022>.

15. Wang, Ming, Generalized Estimating Equations in Longitudinal Data Analysis: A Review and Recent Developments, *Advances in Statistics*, 2014, 303728, 11 pages, 2014. https://doi.org/10.1155/2014/303728

APPENDIX

Table 23 | WoE and IV of risk driver turnover

Category	WoE	IV
[0.15,0.5]	6.58	0.39
[0.5,1.5]	5.39	1.60
[1.5,4]	4.42	1.51
[4,7]	4.00	0.81
[7,10]	3.54	0.26
-	-	$\sum$ IV = 4.57

Source: Simulated data

Table 24 | WoE and IV of risk driver debt ratio

Category	WoE	IV
[0,1]	5.92	4.13
[1,1.5]	3.93	0.76
[1.5,3]	2.77	0.24
-	-	$\sum$ IV = 5.13

Source: Simulated data

Table 25 | WoE and IV of risk driver delinquency

Category	WoE	IV
0	5.76	5.15
1	2.34	0.19
-	-	$\sum$ IV = 5.35

Source: Simulated data

Table 26 | Cramer's V between RDs

	Turnover	Debt Ratio	Delinquency	Seniority
Turnover	–	0.009	0.015	0.006
Debt Ratio	–	–	0.026	0.278
Delinquency	–	–	–	0.026

Source: Simulated data