UNIVERSITA' DEGLI STUDI DI PAVIA
Dipartimento di Scienze Economiche e Aziendali
Corso di laurea magistrale in Finanza

From Accuracy to Profitability:

Evaluating Credit Rating Models' Economic Impact

Relatore:
Chiar.mo Prof. Paolo Stefano Giudici

Tesi di laurea di
Nikolaos Papadopoulos
Matr. n.559129

Anno accademico 2024 - 2025

DEDICATION

“This thesis is dedicated to my parents, my sister,
my beloved family Bira, my friends,
and to my precious girl,
who all of them inspired me with their love.”

ACKNOWLEDGEMENTS

While writing this thesis, only moments from the last 2 years were passing through my mind. How the time is flying so fast, and how much we can achieve from it are two questions impossible to be answered by a simple double degree student. For this reason, I feel the need to express my sincere gratitude to some people who, in their own way, contributed to my progress and my ability to deliver this thesis with success.

Primarily, I would like to thank my supervisor, *Mr. Francesco Nisi*, Credit Risk Analyst at the company named Prometeia, based in Bologna, for his invaluable guidance and time that we spent together discussing this master's thesis. His effort to share important information, explain different topics and spend time working on this thesis was definitely something that gave a unique character and image to this analysis.

I am also really thankful to my professors, *Mr. Dimitrios Karlis* and *Mr. Paolo Stefano Giudici*, from both universities (Athens University of Economics and Business and Università degli Studi di Pavia, respectively), that for their long experience in the fields of Statistics and Finance, they managed to be excellent supervisors, guiding and orienting me throughout the last months.

Through these moments, there are also these people who didn’t come in direct contact with my thesis, people who didn’t give me any advice, or even didn't know what I had done so far, and I am really grateful to have them around me, including friends and university classmates.

Lastly, I dedicate this part of the thesis specifically to my parents and sister, who all believed in me and saw me leaving my homeland, also to the family Bira, who supported me in their own beautiful way in every aspect and we had lovely moments together, and lastly to this special and unique girl, Giulia, who, I can say… got more stressed about this thesis than me!

ABSTRACT

Nikolaos Papadopoulos

From Accuracy to Profitability:
Evaluating Credit Rating Models' Economic Impact

September 2025

This study focuses on the economic value of various predictive accuracy metrics in credit rating models. The general logic of the banks is that they rely on regression-based approaches, while more recently, it's common to see the use of machine learning techniques to assess borrower risk. However, the question of whether investing in higher-performing models or not generates countable financial benefits remains underexplored, and most importantly, underanalysed. This paper addresses this gap by examining how more enhanced discriminatory power in models affects not only profitability, but also the lending quality and regulatory capital requirements of the banks.

The analysis identifies three primary transmission channels, that are explained thoroughly in the main text, through which the models' accuracy can influence different economic outcomes given that: (1) improved loan origination reduces defaults and enhances margins by better identifying low-risk borrowers; (2) stronger models mitigate adverse selection, a very vast sector in banking, helping retain creditworthy clients who might be lost to competitors otherwise; and (3) more accurate models, so more higher value metrics and, by extension, risk assessments can reduce Risk-Weighted Assets (RWA), freeing regulatory capital.

In order to address this request in a more direct way, we are using simulation-based methods, generating synthetic loan portfolios (50,000 loans at 3% default and 10,000 prospects at 10% default) and evaluating models across Area Under the Receiver Operating Characteristic curve (AUROC) bands from 65% gradually increasing to 90%. In order to do that, based on references in the bibliography, we are confident to

use different logistic distributions that were applied to mimic predictive scores, and they were calibrated to ensure consistent risk levels. In the end, the results show that defaults among top-approved loans decline sharply with better accuracy models - from nearly 6% at AUROC with 65% accuracy to less than 1% at AUROC with 90%. Proceeding to the adverse selection analysis, we can confirm that stronger models attract and retain more profitable clients. The capital impact is smaller but meaningful, with average RWA reductions of around 8% between lower-and higher-accuracy scenarios.

Lastly, the profitability that was measured from the previous analysis gives further value to the model improvements. By applying a realistic arithmetical example, on a €3.5 billion retail portfolio, each 5-point AUROC percentage increase can generate approximately €0.8-1 million in addition to the annual profit, with relative gains of 5-12% depending on competitive dynamics. These effects can compound over time as new loans are added annually, while the findings show us that even with incremental improvements in model discrimination can yield and generate significant economic returns, reinforcing the strategic importance of continuous model enhancement. Banks, regulators, and model developers at the same time can use these insights to firstly justify investments, then set performance benchmarks, and also better understand the link between model validation metrics and real-world financial outcomes.

SINTESI

Nikolaos Papadopoulos

Accuratezza e Redditività:
L'Impatto Economico dei Modelli di Valutazione del Credito

Settembre 2025

Il presente studio si concentra sul valore economico di diverse misure di accuratezza predittiva nei modelli di valutazione del merito creditizio. La logica generale delle banche si fonda su approcci basati sulla regressione, mentre più recentemente è diventato comune osservare anche l'uso di tecniche di machine learning per la valutazione del rischio dei mutuatari. Tuttavia, la questione se l'investimento in modelli con performance più elevate produca benefici economici misurabili rimane poco esplorata e soprattutto poco analizzata. Questo lavoro affronta tale lacuna esaminando come un maggiore potere discriminante dei modelli influisca non solo sulla redditività, ma anche sulla qualità e sull'accuratezza dei prestiti e sui requisiti di capitale delle banche.

L'analisi individua tre principali canali di trasmissione, spiegati in dettaglio in seguito, attraverso i quali l'accuratezza dei modelli può incidere in modo differente sui risultati economici, considerando che:

(1) una migliore concessione dei prestiti riduce le insolvenze e aumenta i margini di profitto identificando con maggiore precisione i mutuatari a basso rischio;

(2) modelli più robusti attenuano o addirittura riducono il fenomeno della selezione avversa, un aspetto cruciale per le banche, contribuendo a mantenere clienti con elevata affidabilità creditizia che altrimenti potrebbero essere persi a favore della concorrenza;

(3) modelli più accurati significano valutazioni del rischio di qualità superiore e, di conseguenza, possono ridurre le Attività Ponderate per il Rischio (RWA), liberando capitale regolamentare.

Per affrontare questo tema in modo diretto, utilizziamo metodi basati su simulazioni, generando portafogli sintetici di prestiti (50,000 prestiti con un tasso di insolvenza del 3% e 10,000 candidati con un tasso di insolvenza del 10%) e valutando i modelli in vari intervalli di AUROC, dal 65% fino al 90%. Per farlo, sulla base di diverse fonti bibliografiche, ci avvaliamo di distribuzioni logistiche differenziate, applicate per simulare i punteggi predittivi e successivamente calibrate per garantire livelli di rischio coerenti. I risultati mostrano che le insolvenze tra i prestiti approvati con i punteggi più alti diminuiscono drasticamente con modelli più accurati: da quasi il 6% con un AUROC del 65% a meno dell'1% con un AUROC del 90%. Passando all'analisi della selezione avversa, possiamo confermare che i modelli più performanti attraggono e fidelizzano clienti più redditizi. L'impatto sul capitale è minore ma significativo, con riduzioni medie delle RWA di circa l'8% tra le fasce di accuratezza più basse e più alte.

Infine, la redditività misurata dall'analisi conferisce ulteriore valore ai miglioramenti del modello. Applicando un esempio numerico realistico, in un portafoglio retail del valore di 3,5 miliardi di euro, ogni aumento del tasso di AUROC di 5 punti percentuali può generare circa 0,8-1 milione di euro aggiuntivi di profitto annuo, con guadagni relativi del 5-12% a seconda della dinamica competitiva. Tali effetti possono amplificarsi nel tempo con l'aggiunta annuale di nuovi prestiti. Le banche, le autorità di vigilanza e i modellisti possono allo stesso tempo utilizzare queste informazioni per giustificare gli investimenti, definire benchmark di performance e comprendere meglio la connessione tra le metriche dei modelli e i loro reali risultati economici.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1
INTRODUCTION TO CREDIT RISK

1.1. Introduction

The research in credit risk modelling has traditionally focused on developing algorithms and models that enhance the predictive accuracy of credit risk assessments. Numerous studies have evaluated the performance of various regression-based models and, more recently, machine learning (ML) approaches using historical loan repayment data and so-called *“alternative”* data sources. However, to the best of our knowledge, until today, little research has been dedicated to assessing the economic impact of improved predictive accuracy. In other words, given the costs associated with developing, implementing, and maintaining rating models, why should a bank invest in more predictive models? What tangible benefits should such improvements yield?

Answering these questions is not only relevant for CEOs, CROs and model developers but also for model validators and regulators, who - by gaining a clearer understanding of the cost-benefit trade-offs - can establish more informed expectations regarding minimum performance thresholds for model approval and better define what constitutes a *"significant"* improvement in predictive accuracy for validation purposes.

The objective of this work is to shed light on the economic implications of enhanced rating model accuracy. First, we identify the key transmission channels through which improvements in model performance translate into economic impact. Then, we attempt to quantify this impact of these channels using simulation approaches.

A quick overview of the paper is organised as follows: Section 2 presents a review of the relevant literature that is necessary for the rest of the analysis; Section 3 provides a comprehensive overview of the channels through which improvements in the discriminatory power of rating models affect the profitability of credit institutions; Section 4 describes the simulation framework adopted in this study, while Section 5 presents the results of the impact analyses on the income statement and risk-weighted assets (RWA). Finally, Section 6 summarises the findings, and Section 7 concludes.

1.2. Background and Motivation

Credit risk modelling plays since ever a crucial role in the decision-making processes of financial institutions, and more particularly of banks. These institutions rely greatly on accurate predictions of the likelihood of default, the probability of repayment, as well as other important financial variables when determining the terms of loans and setting aside adequate provisions for potential losses. As the complexity and scale of global financial markets constantly continue to grow, the traditional credit risk models face increasing pressure to provide more precise and reliable estimates. Some early methods, often based on logistic regression or other statistical techniques, have drastically evolved over time to incorporate advanced machine learning algorithms, which offer substantial improvements in prediction accuracy.

However, despite some significant advancements in model development, the economic implications of these improvements have not been extensively studied yet. While the use of more sophisticated models can be expected to yield more accurate predictions, the direct financial benefits to banks - such as increased profitability, reduced capital requirements, and better management of default risk - remain somewhat speculative. In particular, the economic trade-offs of investing in model development, especially when considering the associated costs, are often underexplored.

As banks face mounting pressure to meet regulatory capital requirements and demonstrate sound risk management practices, understanding the economic benefits of improved credit risk models becomes crucial. In this context, the motivation for this research stems from the need to bridge the gap between technical model development and the economic realities faced by banks. By quantifying the financial impact of higher predictive accuracy, we can provide a clearer justification for the investment in more accurate and robust credit risk models, offering both theoretical and practical insights.

1.3. Research Questions and Scope

The primary research question guiding this study is: How do improvements in the predictive accuracy of credit risk models translate into tangible economic benefits for banks? Subordinate to this question are several sub-questions, including:

- What are the key channels through which predictive accuracy influences a bank's profitability?
- How can improvements in model performance lead to more efficient capital utilisation and risk management?
- What is the magnitude of the economic impact of different levels of accuracy in credit risk models, and how does this impact vary across different types of loan portfolios?
- How do adverse selection effects, stemming from improved model accuracy, influence the bank’s risk exposure and profit margins?

Another important note is that this study primarily focuses on retail portfolios, such as mortgage portfolios, where credit risk modelling plays a significant role in managing the risk of default and optimising loan origination strategies. While the focus is on a small, low-risk portfolio, the principles and findings are applicable to larger, more diverse portfolios as well. The results presented will be based on a set of conservative assumptions, reflecting a moderate, yet realistic, view of a bank’s portfolio and its potential for growth and profitability. The scope of this research is, therefore, twofold. On one hand, it provides a conceptual understanding of the economic channels through which improved model accuracy can benefit financial institutions. On the other hand, it attempts to quantify the specific economic effects, providing banks with evidence-based insights into the potential returns from investing in more accurate credit rating models.

CHAPTER 2

LITERATURE REVIEW

The literature review of this thesis focuses on various sources where research has flourished over the last decades, where one of the most substantial parts of modern research has targeted specifically the comparison of the discriminatory power of different credit scoring models that are used widely to predict default probabilities. Additionally to that, some traditional approaches, i.e. logistic regression, have long been the fundamentals in credit risk due to their easy interpretability and regulatory acceptance. However, keeping in mind the growing availability of all sorts of data and also the computational resources of today, studies that came up more recently, have increasingly investigated and explored the use of advanced machine learning (ML) techniques and algorithms, such as the Support Vector Machines (SVM), the Random Forests and the Deep Neural Networks (DNN), in an attempt to improve the predictive performance.

For example, Brown and Mues (2012) have compared, under different levels of class imbalance, various models, which is a typical feature of credit scoring problems. Moreover, Baesens et al. (2015) completed in time an extensive benchmarking of classification algorithms that were used in the credit scoring domain, while some years later, Fuster et al. (2022) compared scoring models with different machine learning approaches, so that they can evaluate their differences in terms of both accuracy and fairness with a more stable approach.

In addition to methodological innovations, researchers have also investigated the contribution of new, non-traditional data sources to model performance. Berg et al. (2018) examine the predictive power of "digital footprint" variables, Giudici et. Al. (2019) augment traditional credit scoring methods with “alternative data” that consists of centrality measures derived from similarity networks among borrowers, Gal et al. (2019) explore the non-traditional sources such as digital applications on mobile phones and e-commerce platform data.

These studies collectively show that both the choice of modelling technique and the incorporation of novel data sources can significantly enhance the discriminatory power of credit scoring models. This always-expanding body of literature provides strong empirical support for the continued exploration and adoption of advanced analytics and alternative data in the credit risk domain. However, while considerable attention has been devoted to enhancing model performance, fewer studies have explicitly examined the monetary consequences of improved discriminatory power, and, for this reason, the economic impact of improved model accuracy in credit rating systems remains a relatively underexplored area in the academic literature.

One of the earliest contributions in this area is Jankowitsch et al. (2005), who proposed a framework for assessing the economic value of internal rating systems by linking model performance to capital requirements and pricing strategies. To the best of our knowledge, the study by Jankowitsch et al. represents the only attempt in the academic literature to jointly consider the impact of an increase in the discriminatory power of credit rating models on both Risk-Weighted Assets (RWA) and the Profit and Loss statement of the bank. Their simulation-based approach explicitly models how enhanced predictive performance can simultaneously influence credit risk capital and the bank's income generation through better credit allocation. This dual focus makes their contribution particularly relevant for understanding the broader economic value of rating model improvements.

2.1. Risk-Weighted Assets and Capital Requirements

2.1.1. Introduction - RWA

Risk-Weighted Assets (RWAs) are a fundamental element of banking regulation, designed to ensure that banks hold sufficient capital to safeguard against potential losses. In their 2012 IMF Working Paper, Sofiya Avramova and Vanessa Le Lesle provide a thorough examination of RWAs and their implications for capital requirements, revealing significant variability across banks and regulatory frameworks (Avramova & Le Lesle, 2012). This variability raises concerns about the consistency and effectiveness of capital adequacy measures.

2.1.2. Capital Requirements and Their Importance

Capital requirements refer to regulatory mandates that require banks to maintain a minimum amount of capital relative to their risk-weighted assets. Avramova and Le Lesle (2012) explain that these requirements are designed to ensure that banks can absorb losses and remain solvent during periods of financial stress. Typically, capital is measured as a percentage of RWAs. For example, under the Basel III framework, banks must maintain a minimum Common Equity Tier 1 (CET1) capital ratio of 4.5%, a Tier 1 capital ratio of 6%, and a total capital ratio of 8% relative to RWAs (Avramova & Le Lesle, 2012).

The authors emphasise that these capital buffers are essential for promoting financial stability by enhancing banks' resilience to shocks and maintaining confidence in the banking system. Adequate capital requirements reduce the likelihood of bank failures and minimise the need for taxpayer-funded bailouts.

2.1.3. Definition and Calculation of Risk-Weighted Assets

RWAs measure the risk exposure of a bank's assets, adjusting the value of each asset according to its perceived credit, market, and operational risk. Avramova and Le Lesle (2012) discuss two primary approaches to RWA calculation permitted under the Basel framework: the Standardized Approach and the Internal Ratings-Based (from now on IRB) Approach. The Standardized Approach assigns predetermined risk weights to asset categories, whereas the IRB Approach allows banks to estimate risk parameters using their internal models. While the IRB Approach provides more risk sensitivity and theoretically leads to more accurate capital requirements, Avramova and Le Lesle (2012) highlight that it also introduces substantial variability. Different banks may adopt diverse methodologies and assumptions, leading to discrepancies in RWAs even for similar asset exposures.

A key finding in Avramova and Le Lesle's study is the considerable variation in RWA estimates across banks, particularly those employing the IRB Approach. Differences in assumptions related to default probabilities, loss given default, and other risk factors cause inconsistencies in RWA measurements (Avramova & Le Lesle, 2012). This variability undermines the comparability of capital ratios across institutions and countries. The practical consequences are significant. Banks with relatively low RWA values may appear more capitalised, enabling them to take on greater risk without commensurate capital buffers. Conversely, banks reporting higher RWAs face stricter capital requirements, which could constrain lending activities and affect their competitive position (Avramova & Le Lesle, 2012).

2.1.4. Policy Recommendations

To mitigate these challenges, Avramova and Le Lesle (2012) propose several policy measures:

1. **Enhanced Disclosure:** Banks should increase transparency regarding their RWA calculation methodologies to facilitate better comparability and market discipline.
2. **Supervisory Review:** Regulators need to rigorously assess banks' internal models and risk assumptions to ensure consistency with regulatory standards.
3. **Standardization:** Introducing more standardized calculation approaches could reduce variability and improve the reliability of capital ratios.
4. **Stress Testing:** Regular stress tests can help evaluate the robustness of capital buffers and the accuracy of RWA estimations under adverse conditions.

Implementing these recommendations would strengthen the effectiveness of capital requirements and contribute to a more resilient banking sector.

In summary, the IMF Working Paper by Avramova and Le Lesle (2012) underscores the critical role of capital requirements and RWAs in banking regulation while highlighting significant inconsistencies in RWA calculations. Addressing these inconsistencies through enhanced transparency, stricter supervisory oversight, standardisation, and stress testing is necessary to maintain the credibility of capital adequacy measures and ensure financial stability.

Nevertheless, this pioneering work still presents several limitations. Specifically, on the RWA side, the impact is assessed under the assumption that a Bank can define its rating scale from scratch. In practice, this is generally not feasible due to IT and business constraints that require IRB Banks to maintain stable rating scales over time. Typically, only new Banks or those transitioning from the standardised approach to the IRB approach have the opportunity to redesign their rating scales to maximise RWA benefits from more predictive models. Therefore, it is essential to evaluate the effect of improved model accuracy on capital requirements even under a fixed rating scale.

Furthermore, the income-generating impact of higher model accuracy is simulated based on two strong assumptions: (1) that the true individual Probability of Default (from now on PD) distribution of borrowers is known - an unverifiable and widely debated assumption - and (2) that borrowers decide whether to accept or reject loan offers based on the difference between their perceived PD and the interest rate offered by the bank, which is itself directly derived from the bank's estimated PD.

Finally, the accuracy of the rating models was measured using the variance of the deviations between predicted probabilities of default and “observed” probabilities of default, making it difficult to translate the results into real-world application contexts. In practice, Banks typically use metrics such as AUROC to evaluate models' accuracy.

2.2. Logistic Distribution

2.2.1. Using the Logistic Distribution in Consumer Credit Scoring

In general, the prediction of default risk has long been one of the most important challenges faced by banks and other lending institutions. When an applicant requests a loan, the lender must assess not only the potential return, but also the likelihood that the borrower will meet their repayment obligations as it is agreed. In recent decades, statistical and econometric tools have played a central role in this assessment, where, among these tools, logistic regression, which is built on the mathematical foundations

of the logistic distribution, has emerged as one of the most widely adopted methods for estimating the probability of default (PD) in credit scoring systems.
The study by Costa e Silva et al. provides a clear illustration of how this approach can be implemented in practice. Their work focuses on data from a Portuguese bank, and more specifically on consumer loans. In short, by applying the logistic regression, they were able to both identify the borrower characteristics most strongly associated with default and create a model that achieved nearly 90% correct classification of cases. So obviously, this makes their research a valuable case study in the intersection between statistical theory and practical risk management.

2.2.2. The Logistic Distribution and Its Role in Credit Scoring

Based on Costa e Silva, the logistic distribution is a continuous probability distribution characterised by its S-shaped cumulative distribution function. Its main appeal in binary outcome modelling lies in how it maps any real-valued input into a probability between zero and one. In logistic regression, a linear combination of predictor variables is transformed through the logistic function:

$$p = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\ldots+\beta_k x_k)}} \tag{2.2.2.1}$$

Here, p is the predicted probability of the event occurring, in this case, the probability of loan default. The parameters are estimated from the data using **maximum likelihood estimation**. The exponential structure ensures that the output is always bounded, avoiding impossible probabilities outside the 0-1 range.

In a credit scoring context, the "event" is usually a default within a certain period. Each predictor, such as a borrower’s income, age, loan characteristics, or past credit behaviour, influences the *log-odds* of default. The advantage is that each coefficient has a clear interpretation: it represents the change in the log-odds of default for a one-unit change in the corresponding variable, holding others constant. This transparency makes logistic regression not only a powerful predictive tool but also one that satisfies regulatory demands for explainability.

2.2.3. Data and Variables in the Portuguese Bank Study

The authors of the study analysed data from 2,577 consumer loan applicants. The dataset included both financial and demographic variables, allowing for a multidimensional view of credit risk. Several predictors stood out in their analysis, like:

- Loan spread – the difference between the loan's interest rate and a benchmark rate, reflecting the cost of borrowing.
- Loan term – the duration of the repayment period.
- Borrower's age – providing insight into life stage and potential financial stability.
- Number of credit cards – potentially a signal of credit experience and discipline.
- Salary institution relationship – whether the borrower's salary was paid into the lending bank, which can strengthen repayment control.
- Income tax bracket – a proxy for income level, with particular attention to those in the lowest bracket.

These variables were not chosen arbitrarily as they align with both financial theory and practical lending experience, where income stability, indebtedness, and institutional ties often influence repayment behaviour.

2.2.4. Model Estimation and Interpretation

The logistic regression model fitted by the authors produced results consistent with economic intuition:

- Higher loan spreads were associated with higher default probabilities, possibly reflecting either higher perceived risk by the bank (and thus higher interest rates) or heavier repayment burdens for the borrower.
- Longer loan terms also increased default likelihood, which is intuitive since a longer repayment period increases exposure to income shocks or economic downturns.

- Older borrowers were found to have higher odds of default, which might reflect a complex mix of factors such as nearing retirement or carrying accumulated debts.
- Interestingly, having more credit cards was associated with *lower* default probability, possibly because multiple active credit lines indicate financial discipline or a higher credit score.
- Borrowers whose salaries were deposited directly with the bank were less likely to default, perhaps because this facilitates automatic payments and signals a closer banking relationship.
- Belonging to the lowest income tax bracket markedly increased default risk, which aligns with the general link between low income and greater vulnerability to financial stress.

The model achieved a classification accuracy of 89.79%, meaning that it correctly distinguished between defaulters and non-defaulters in nearly nine out of ten cases. This level of performance is notable in the field of credit risk modelling, where prediction is inherently probabilistic and perfect classification is unattainable in every case.

2.2.5. From Logistic Regression to Credit Scoring

The connection between logistic regression and credit scoring is straightforward yet powerful. Once the model coefficients are estimated, the logistic equation can be applied to any applicant's data to generate a Probability of Default (PD). Lenders can then establish score thresholds to categorise applicants into risk groups, such as "low risk," "moderate risk," or "high risk."

Moreover, because the coefficients have direct interpretations, loan officers and risk managers can explain why a given applicant's score is high or low. For example, if an applicant's PD is high due to a long loan term and low income bracket, these factors can be directly communicated and, in some cases, adjusted through loan restructuring.

2.2.6. Comments for Future Expansions

Costa e Silva et al.'s model demonstrates the enduring relevance of logistic regression in the age of more complex machine learning algorithms. While decision trees, gradient boosting, and neural networks have gained attention, logistic regression remains a preferred choice in banking due to its transparency, computational efficiency, and compliance with regulatory frameworks such as the Basel Accords.

The study by Costa e Silva et al. offers a compelling example of how the logistic distribution and logistic regression framework can be applied to credit risk assessment. By combining well-chosen predictors with a robust statistical method, they produced a model that is both interpretable and highly accurate. The findings reinforce why logistic regression remains a mainstay in credit scoring: it balances predictive strength with transparency, allows direct estimation of default probabilities, and can be validated and explained in ways that satisfy both internal decision-makers and external regulators.

In a financial landscape where lending decisions carry significant consequences for both banks and borrowers, the ability to accurately and transparently assess credit risk remains indispensable. The approach demonstrated in this research shows how statistical theory, grounded in the logistic distribution, continues to deliver practical, high-impact solutions for one of the most persistent challenges in banking.

2.2.7. Logistic Regression in Credit Scoring

One of the thesis's purposes is also to explore in depth how logistic regression can be applied in practice to generate such scores, thus enabling banks to assess and prioritise credit applicants more effectively. As discussed before, the dataset used originates from a publicly available source containing records for over 3,000 banking customers, each described by fourteen independent variables. These include demographic factors, financial attributes, and borrowing details. The principal aim is to estimate a logit model, a variant of logistic regression, capable of assigning a risk score to each applicant, where higher scores correspond to higher estimated likelihoods of default.

The analytical process begins with preliminary treatment of the raw data, including descriptive diagnostics such as skewness and kurtosis to characterise the distributions of each input variable. To mitigate the influence of extreme outliers, a winsorization procedure (capping values at selected percentiles, here at 0.5%) was applied. This data cleansing step helps ensure that exceptionally large or small observations do not disproportionately distort coefficient estimates within the regression.

Following data cleanup, an initial logistic regression is conducted on the winsorized dataset. The output presents estimated coefficients for each independent variable, along with standard errors, t-ratios (or Z-statistics) and corresponding p-values. These metrics allow the identification of variables whose association with the default outcome is not statistically significant at the established threshold (5% significance level). In this analysis, three predictors, loan term, loan price, and savings, are found to produce p-values above this cutoff, indicating they do not contribute explanatory power in explaining default risk in this dataset.

These three variables are then excluded in a refined model specification. A second logistic regression, now including only the remaining significant predictors, is estimated to derive more efficient coefficient estimates. An overall likelihood-ratio (from now on LR) test is performed to gauge the goodness-of-fit of the model. The result (a p-value of approximately 97.8%) suggests that adding back the excluded variables would not materially improve the model's explanatory capacity. In other words, the more streamlined set of predictors (excluding loan term, price, and savings) yields a model that is robust and statistically justified.

At its core, this study highlights how the logistic distribution serves as a natural and effective foundation for credit scoring. The logistic function converts the linear combination of explanatory variables into a sigmoid-shaped curve, mapping any real-valued score into a probability in the interval (0, 1). This mapping respects the theoretical requirement that probabilities cannot exceed 100%, while preserving interpretability: the log-odds of default is expressed linearly in the predictors, and exponentiation of these log-odds yields direct PD estimates. As a result, the logistic model excels at distinguishing between low-risk and high-risk applicants and ranking them accordingly.

By harnessing logistic regression, the credit scoring process becomes transparent and quantifiable: each predictor's coefficient indicates how a unit change influences the log-odds of default, enabling risk managers to interpret model output in business terms. The statistical tests (such as Wald tests for individual predictors and the LR test for model fit) support model validation, ensuring that each included variable adds meaningful predictive power. Winsorization and other data preprocessing steps further strengthen model reliability by reducing the impact of anomalous data points.

In conclusion, this thesis demonstrates that logistic regression, grounded in the logistic distribution, is a sound and practical method for constructing credit risk models. From input preprocessing and significance testing to final score computation and PD mapping, the approach yields a statistically rigorous and interpretable tool. By carefully selecting relevant predictors and mapping linear risk scores through the logistic curve, banks can reliably rank applicants by default risk and make informed lending decisions—all while aligning with established credit risk modelling practices guided by the bounded and smooth nature of the logistic distribution.

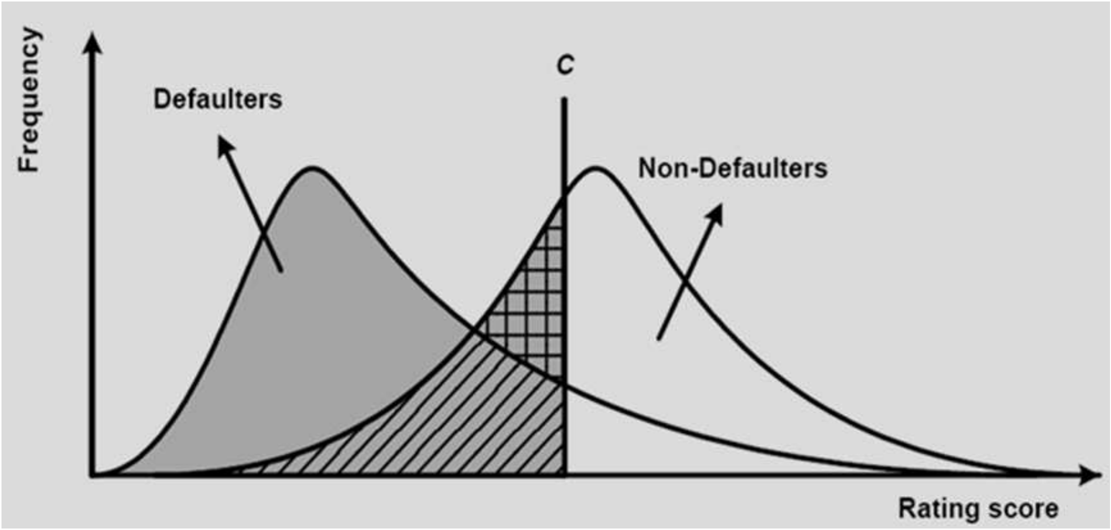
2.3. Metrics' Analysis and AUROC

2.3.1. Classification Procedure

The following analysis, done by Stephen Satchell and Wei Xia (2006), stresses that this issue has been analysed in the **BCBS Working Paper No. 14 (2005)**, which provides a summary of various statistical techniques designed to assess a model's ability to distinguish between outcomes. For this reason, several approaches have been proposed in academic and professional literature, including the Cumulative Accuracy Profile (from now on CAP), the Receiver Operating Characteristic (from now on ROC) curve, Bayesian error rates, the Conditional Information Entropy Ratio (from now on CIER), Kendall's tau, Somers' D, and the Brier score, among others. Among these, the CAP (along with its associated metric, the Accuracy Ratio (from now on AR)) and the ROC curve, summarised by the AUROC, are the most widely used till date.

Furthermore, they consider it important to mention that, unlike certain metrics which ignore sample size and can be significantly influenced by statistical variability, the CAP and ROC-based measures in fact consider the sample size of defaulters. As a result, they are more robust and reliable for comparing the effectiveness of different rating models.

Figure 2.1: Distribution of Rating Scores



(Stephen Satchell and Wei Xia (2006))

To assess the likelihood of borrowers defaulting or not, we often rely on rating scores. Ideally, a perfect model would show a clear separation between the score distributions of defaulters and non-defaulters. However, in practice, these distributions usually overlap, as it is illustrated in Figure 1.

In Figure 1, the horizontal axis represents the rating score, while the vertical axis shows the frequency. The score marked as **C** serves as a decision threshold. Borrowers scoring below **C** are flagged as potential defaulters, and those scoring above are treated as non-defaulters.

With this setup, there are four possible classification outcomes:

1. A borrower scores below **C** and defaults → this is a **correct classification**.
2. A borrower scores below **C** but doesn’t default → this is a **false positive**.
3. A borrower scores above **C** and doesn't default → another **correct classification**.

4. A borrower scores above **C** but defaults → a **false negative**.

From here, we define the **hit rate**, HR(C), which tells us how effectively the model identifies true defaulters using cut-off **C**:

$$HR(C) = \frac{H(C)}{N} \tag{2.3.1.1}$$

Here, H(C) is the number of actual defaulters correctly identified, and N_D is the total number of defaulters in the dataset. The **false alarm rate**, FAR(C), captures the proportion of non-defaulters wrongly flagged as defaulters:

$$FAR(C) = \frac{F(C)}{N_{ND}} \tag{2.3.1.2}$$

In this formula, F(C) is the number of non-defaulters incorrectly predicted to default, and N_{ND} is the total number of non-defaulters. Visually, in Figure 1:

- H(C) corresponds to the shaded area on the left of **C** under the defaulters' curve.
- F(C) corresponds to the hatched area on the left of **C** under the non-defaulters' curve.

2.3.2. Area Under the ROC Curve (AUROC)

To evaluate the model comprehensively, we use the **Receiver Operating Characteristic (ROC) curve**. This curve plots HR(C) against FAR(C) for every possible value of **C**. The ROC curve helps judge how well the model distinguishes between the two groups. The AUROC, denoted as **A**, summarises this performance:

$$A = \int_{0}^{1} HR(FAR)d(FAR) \tag{2.3.2.1}$$

A perfect model has an AUROC of 1, indicating flawless separation of defaulters and non-defaulters. On the other hand, a model with no predictive power will have an AUROC around 0.5, equivalent to random guessing. Most useful models in real-world scenarios tend to fall between these extremes.

However, in practice, this integral cannot be solved and calculated analytically because we don’t have a continuous function describing HR(FAR). Instead, what we have are discrete empirical data points, derived from the cumulative hit rates and false alarm rates calculated across rating buckets sorted by predicted default probability. These elements form the basis for plotting the empirical ROC curve, and to approximate the area under this curve, we apply the trapezoid rule, a common numerical integration technique. In short, the trapezoid method works by partitioning the ROC curve into a series of adjacent trapezoids, each spanning between two consecutive points on the curve. It then sums the areas of these trapezoids to estimate the total area under the curve. This approach remains a good selection for the ROC analysis because:

1. **It handles non-linearity between points:** The ROC curve is not guaranteed to be linear between points, and the trapezoid rule provides a more accurate approximation than simply summing rectangles or assuming a uniform step size.
2. **It requires minimal assumptions:** Unlike other integration techniques that assume smooth or differentiable functions, the trapezoid rule works reliably with the step-wise empirical data derived from model predictions.
3. **It’s computationally efficient:** Especially in large datasets common in credit risk modelling, the trapezoid method balances precision and speed, making it ideal for practical applications.

Considering the above analysis, done by Stephen Satchell and Wei Xia (2006), we believe a more realistic approach assumes a competitive market environment in which each borrower asking for a loan accepts the offer from the bank which provides the most favourable conditions (i.e., the lowest interest rate). In such a setting, Banks with more accurate rating models, in terms of AUROC, are better positioned to offer more competitive terms to truly low-risk clients, thereby reducing the likelihood that these clients receive (and accept) better offers elsewhere.

Stein (2004) connected ROC analysis with loan pricing. His work is especially valuable in formalising the idea that discriminatory power directly affects profit through better risk segmentation. Nonetheless, Stein himself highlights the limited generalizability of his results, cautioning that “it is not possible to draw specific inferences about the value of a particular power differential [...] since the value of one model versus another is sensitive to both the economic conditions under which the models would be used, and the shapes of the two power curves". This underscores the need for more robust and generalizable simulation frameworks.

Bloechlinger and Leippold (2005) quantify the economic benefit of more powerful credit scoring systems by analysing reductions in expected loss and improvements in profit margins. They show that even marginal improvements in model accuracy can lead to notable financial gains for lenders, especially in the lower end of the score distribution.

More recently, Alonso-Robisco and Carbó (2022) investigated the capital savings potential from adopting machine learning models in IRB frameworks. They find that ML-based models can lead to significant RWA reductions. However, their focus is predominantly on the form of the model (e.g., XGBoost vs. logistic regression) rather than on performance metrics like AUROC per se. This distinction is crucial: while model form and performance are related, conflating the two can obscure the real driver of economic benefit—the ability of a model to correctly rank-order risk.

Zhou and Van Vuuren (2024) add to the discussion by examining how deteriorations in discriminatory power influence expected losses and capital adequacy. Their empirical evidence shows that even modest declines in AUROC can substantially increase expected losses and required capital, reinforcing the economic relevance of maintaining high model accuracy. However, they do not attempt to build a holistic cost-benefit model that includes lending profits, capital charges, and borrower behaviour.

These studies, while valuable, present (at least one of) the following three main weaknesses that we aim to address:

1. **Partial economic assessment**: Most studies isolate one dimension of the model's economic impact - either capital savings (RWA) or improved lending outcomes - without providing a comprehensive evaluation of the total potential benefit.

2. **Conflation of model form and model performance**: Several works, especially those incorporating machine learning, focus on the functional form of the model rather than isolating the impact of discriminatory accuracy (e.g., AUROC or Gini). We argue that the true economic value should be assessed in terms of ranking performance, independent of model structure.
3. **Lack of generalizability**: As noted by Stein (2004), many studies rely on specific datasets, which limits the external validity of their results.

In contrast, our research proposes a simple, yet comprehensive and generalizable simulation-based framework aimed at providing a conservative estimate of the economic gain achievable through improvements in rating models' discriminatory power.

It incorporates both the lending-side impacts (e.g., improved borrower selection and pricing) and capital-side impacts (e.g., RWA optimisation), and considers heterogeneous bank types (e.g., new entrants vs. established IRB users). This framework can be easily tailored to provide more accurate estimates for specific portfolios. Crucially, it remains independent of the underlying model specification and focuses solely on the impact of changes in AUROC. This allows for more robust and generalizable conclusions regarding the financial value of improving model accuracy.

CHAPTER 3

TRANSMISSION CHANNELS' OVERVIEW

In order to obtain a quantification of the economic impact of an increased accuracy of rating models, it is first necessary to identify the mechanisms through which an increased (decreased) predictive ability of rating models turns into an economic gain (loss) for the bank. The main drivers are the following:

1. **Impact on Expected Losses - Lending**: The first and most immediate impact of increased accuracy of rating systems on bank profitability is related to the loan origination phase. Using rating models that are more predictive at the origination stage makes it possible to optimise the quality of the loan portfolio, granting fewer loans to counterparties that will go into default and at the same time increasing margins on performing customers. This channel represents the effect of the model's discriminatory power per se, that is, the impact of a given *absolute* level of model performance (e.g. AUROC) on the Bank's ability to grant credit to more or less creditworthy borrowers.
2. **Impact on Expected Losses – Adverse Selection Effect:** the lending effect (point 1 above) is further amplified in the pricing phase. In fact, what happens in practice is that Banks accept a certain subset of the loan requests they receive, but those, among these "accepted requests", who are "overpriced" (i.e. customers with good risk profile to whom the Bank charges high interest rates because its model suggests they have high estimated PDs) can turn to another bank that offers terms more consistent with their risk profile. If even some of these good clients change banks, the return of the Bank's portfolio will be correspondingly lower. Unlike the previous transmission channel, this one represents the *relative* impact of a model's discriminatory power on the quality of a Bank's credit portfolio. In other words, it is not sufficient for a model to perform reasonably well in absolute terms to ensure an equally reasonable portfolio quality if competing Banks have better-performing models and can therefore offer customers terms more closely aligned with their risk profiles. This effect will become increasingly pronounced as markets

become more efficient and frictionless - thanks to fully digital credit offer aggregation and disbursement systems, and increased competition from fintechs and new banks - making it easier for customers to switch between banks in search of the best contractual terms.

Example: Imagine two banks assessing the same loan applicant. Bank A uses a highly accurate model and estimates a PD of 0.5%, offering a 3% interest rate. Bank B, with a less accurate model, estimates PD at 3% and offers a 6% interest rate. The applicant obviously chooses Bank A. Therefore, bank B loses a low-risk borrower, leading to a portfolio skewed toward higher-risk clients, raising its expected losses.

3. **Impact on RWA**: models with higher predictive power tend to have benefits in terms of RWA (lower capital required)
    a. In the usual case in which the Bank has a rating scale already in place, this is not a strict mathematical constraint, but rather a general trend resulting from two main factors:
        i. The first is the shape of the capital requirement function for PD, which is concave - meaning it requires low amounts of capital for both very low and very high PD values. In fact, the capital requirement is zero when PD = 0% as well as when PD = 100%.
        ii. The second factor is the expected behaviour of models as their discriminatory power increases. To get an intuition for this, one can imagine the behavior of a perfect or ideal model: such a model, if it existed, would always assign a PD = 0% to all clients who will not default in the next 12 months (default flag = 0), and a PD = 100% to all clients who will default (default flag = 1). As a result, the RWA capital requirement would be zero. Models with increasing discriminatory power (i.e., models approaching perfection), being increasingly capable of distinguishing between good and bad clients, will tend to concentrate their predictions at the extremes of the distribution. This, in turn, leads to a reduction in the required RWA capital.

This effect is present even when the bank uses the same rating scale.

b. Nevertheless, in the (unusual) case of financial institutions that can define a new rating scale from scratch (e.g., because they are new Banks developing an IRB rating system for the first time, or Banks moving from the Standardized Approach for capital requirement calculation to the IRB Approach), this effect becomes even more pronounced. In fact, financial institutions with more predictive models will have the ability to define more granular rating scales. Due to Jensen's inequality (as mentioned, the capital requirement curve for PD is concave), this leads to a lower capital requirement, all other conditions being equal.

Hypothetical Illustration: A portfolio with an average PD of 2% may require a certain capital buffer. If a better model reclassifies low-risk clients with PD < 0.5%, and these clients are assigned appropriately low capital weights, the bank could see a 10–20% reduction in RWA, improving its capital efficiency.

4. **Impact on Expected Losses - Monitoring**: especially in the case of portfolios with large credit lines (e.g. Corporate / Large Corporate portfolios, but also Retail - Cards): having a more predictive model available makes it possible to focus monitoring activities on the riskiest customers and to timely reduce exposure to customers who will default. This exposure-reducing effect does not apply to instalment-based portfolios (e.g., mortgages), where the presence of a payment schedule contractually defines the exposure.
5. **Impact on regulatory add-ons**: deficiencies in the predictive capacity of models, even on specific clusters, may generate regulatory capital add-ons when models are inspected by the regulator (e.g. ECB).
6. **Other indirect impacts**: There are also a large number of additional indirect effects of improved rating model accuracy. Among these are: if RWA decreases, the CET1 ratio improves, making the Bank appear more solid; More accurate and conservative models lead to lower variability in RWA, thereby reducing Banks' funding costs and, consequently, increasing economic

profit. If the models are more effective, end users will also use them more efficiently and more frequently (with fewer overrides), etc.

As one can easily infer, the accuracy of rating systems has multiple consequences/impacts, thus making an overall assessment of the impact of improved predictive ability of models very complex.

In this paper, we will therefore focus on the **first 3 aspects only**, because they are valid for all portfolios (unlike 4, which does not apply to instalment portfolios) and easier to estimate (unlike 5, which would require particularly confidential information which is available to the regulator only, and 6). The objective of the study is therefore to obtain a **conservative estimate** of the expected economic benefits of increasing the performance of rating systems through transmission channels 1, 2 and 3.

While this study deliberately adopts a conservative scope by focusing only on transmission channels 1-3, it is important to note that these channels represent the most universal and quantitatively tractable impacts of model improvements. They capture the core profit and capital optimisation levers available to any bank, regardless of portfolio type or regulatory idiosyncrasies.

In the following chapter, we operationalise these transmission channels into a simulation framework. This framework allows us to quantify the economic implications of improvements in model discriminatory power, focusing on lending decisions, adverse selection dynamics, and RWA capital savings.

CHAPTER 4

DESCRIPTION OF THE SIMULATION SETTING

Iftho Hara Khanam (2023) comments that in many economies around the globe, commercial banks play an important role in sustaining economic activity by extending credit to individuals and businesses and, accordingly, by enhancing the efficiency of the credit-granting process has become a priority. A decent banking sector that is equipped with well-designed credit evaluation systems can contribute significantly to economic growth, while enabling firms and consumers to access financing, and banks help generate jobs and foster enterprise expansion. Nevertheless, if credit is granted to high-risk borrowers without effective risk controls, the consequences can ripple across financial systems, potentially triggering systemic instability. Historical crises have demonstrated many times in the past that unchecked lending to credit-unworthy customers can erode confidence and destabilise entire economies.

In response to these challenges, financial institutions have increasingly adopted systematic approaches to credit risk management, where, among these, credit scoring, particularly using logistic regression models, has emerged as a central tool. Logistic regression, which relies on the logistic (or sigmoid) function, translates risk-related predictor variables into a score that is then linked to a probability of default (PD) as discussed before. This approach not only supports comparative ranking of borrowers based on risk but also produces probabilities bounded between 0 and 1, making it intuitively interpretable and statistically sound.

To overcome the problem of lack/confidentiality of data and to allow for a higher generalizability of results, we opted for a simulative approach in which we generate two samples:

- **Bank Loans portfolio at Time 0** consisting in 50,000 loans (with equal Loss Given Default (from now on LGD) and Exposure At Default (from now on EAD)) and 3% Default Rate (i.e. 1,500 observations out of 50,000 where flagged with Default Flag = 1 meaning that they will not repay their debt within 12 months, while 48,500 observations out of 50,000 were flagged with

Default Flag = 0 meaning that they represent the good/performing part of the Bank portfolio)

- **List of Prospect Clients** (i.e. Prospect clients who are not already customers of the Bank but could become so since they have applied for a loan which may, or may not, be granted by the Bank). This represents the market of potential borrowers who applied for a loan to the Bank at Time 0. It consists of 10,000 loan applications with a 10% Default Rate. The LGD and EAD of these potential borrowers are constant and equal to the LGD/EAD of the Bank portfolio[1].

For each of these two portfolios for which we assume to know the ground truth (i.e. their default behaviour within the next 12 months), we generate a set of predicted Probabilities of Defaults (PDs). Specifically, we generate 100 distributions of Predicted PDs for each AUROC within each of the following ranges: 64-66, 69-71, 74-76, 79-81, 84-86, 89-91.

These predicted PD distributions were randomly generated by drawing numbers from two different logistic distributions (differing in location parameter) for 'Good' and 'Bad' clients. The closer the location parameters of the two distributions, the lower the expected AUROC; conversely, the further apart the location parameters, the higher the expected AUROC. Distributions were generated until 100 sets were obtained for each AUROC range (64-66, 69-71, 74-76, 79-81, 84-86, 89-91).

*[**START**] TECHNICAL DESCRIPTION OF THE SAMPLING PROCEDURE*

Sample Generation Process for AUROC-Based Simulations

To evaluate the models' performance across varying levels of discriminatory power, we simulated data reflecting different ranges of AUROC (Area Under the ROC Curve). The way that this procedure was structured is the follows.

---

[1] *Undoubtedly, improvements in the predictive capabilities of LGD and EAD/CCF models also have a significant expected economic impact; however, such an assessment falls outside the scope of the present work.*

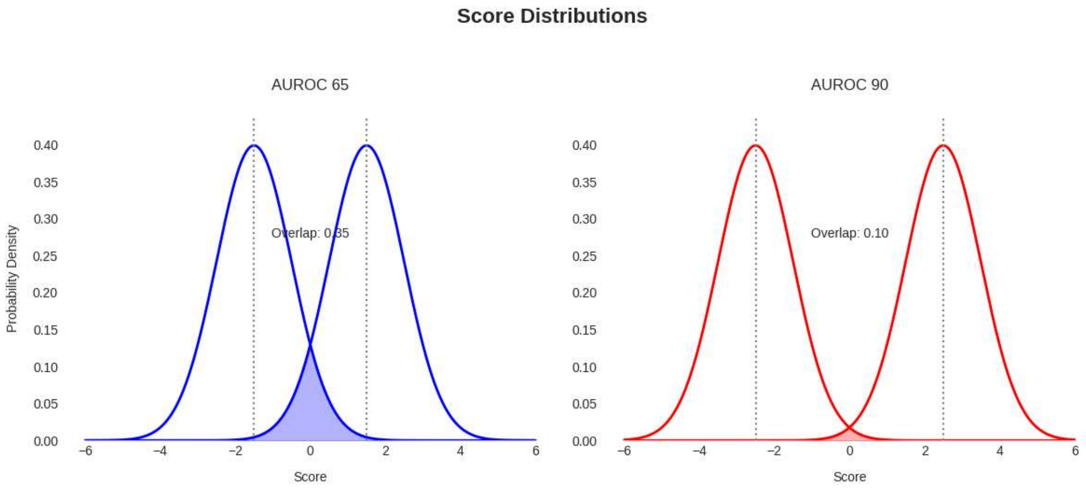
- **Step 1: Defining Population Structure**

  This 50,000 population represent the bank’s loan portfolio at time = 0. We simulated a binary classification scenario involving:
  - **48,500 non-default observations** (Flag = 0)
  - **1,500 default observations** (Flag = 1)

- **Step 2: Setting Logistic Score Parameters**

  To control the AUROC, we specified different location parameters (loc0, loc1) for the logistic distributions used to generate scores:
  - loc0 controls the score distribution for non-defaults.
  - loc1 controls the score distribution for defaults.
  - The separation between loc0 and loc1 affects the AUROC.

Figure 4.1: Impact of AUROC on Score Separation



This comparison shows us how well a scoring system can tell apart two groups (like good vs. risky borrowers). Both pictures show two bell curves representing the scores for each group. The main difference? How much the curves overlap, which tells us how often the system might mix them up.

Left Picture ("AUROC 65"):

- The two curves are somewhat separated (peaks at -1.5 and 1.5).

    - They overlap by **35%** (blue area), meaning there’s a decent chance some risky borrowers will get scores similar to good ones (and vice versa).

  Right Picture ("AUROC 90"):

    - The curves are much farther apart (peaks at -2.5 and 2.5).
    - Only **10% overlap** (red area) - the system rarely mixes up the two groups.

- **Step 3: Simulating Predicted Scores (Logistic Distribution)**
  For each AUROC range (64-66%, 69-71%, 74-76%, 79-81%, 84-86%, 89–91%), we repeatedly drew samples from the **logistic distribution** using the specified loc0 and loc1 values:
    - Non-default scores were drawn from Logistic(loc=loc0, scale=√10/π).
    - Default scores were drawn from Logistic(loc=loc1, scale=√10/π).
    - The predicted probability of default (PD_pred) was then computed via the sigmoid function:

$$PD_{pred} = \frac{1}{1+e^{-score_{pred}}} \tag{4.1}$$

- **Step 4: Assign Rating Classes**
  We assume that the bank has a 10-class rating scale already in place. Each simulated observation was assigned to a rating class (1 to 10) based on thresholds applied to the PD_pred. These thresholds are taken from credit rating breakpoints used for EU Banks Pillar 3 reporting (e.g., <0.0015, 0.0015 - 0.0025, ..., >0.3).

- **Step 5: Filter by AUROC Range**
  For each simulation:
    - We grouped observations by rating class.
    - We computed cumulative true and false positive rates.
    - The AUROC was estimated using the trapezoidal rule on the ROC curve.

  - Simulations were accepted only if their AUROC fell within one of the target ranges (i.e., 64-66%, 69-71%, 74-76%, 79-81%, 84-86%, 89-91%). This process was repeated until 100 valid simulations were collected for each AUROC band (or until a maximum number of attempts was reached).

- **Step 6: Calibrating PD Predictions**

  This is needed to guarantee that all the simulated PD distributions point at the same average risk level (regardless of their AUROC level) so ensuring fair comparisons. Since the raw PD_pred values did not necessarily match the desired portfolio mean PD (3%), we applied a logit-shift calibration on the logit scores[2]:
  - First, convert PD_pred to log-odds (logit scores).
  - Then, get the optimal shift that brings the average PD to the 3% target.
  - Finally, transform back to probability space using the sigmoid function.

- **Step 7: Saving and Storing Results**

  For each valid simulation:
  - Both the original and calibrated PD_pred vectors were stored.
  - Calibrated vectors ensured consistent average portfolio PD across AUROC bands.
  - All results were stored in a comprehensive dataframe (simulation_df), with each column representing the calibrated PD predictions from one simulation.

[__END__] TECHNICAL DESCRIPTION OF THE SAMPLING PROCEDURE

We used these simulated data to obtain a quantification of the economic impact of models' accuracy for each of the transmission channels described in Section 3.

---

[2] *Since applying an additive shift directly to the PDs could have resulted in values falling outside their valid range [0,1], we chose to transform the PDs into logistic credit scores. We then applied a shift to the scores such that, after transforming the shifted scores back into PDs, the average PD matched the target PD.*

CHAPTER 5
QUANTIFICATION OF THE IMPACT

5.1. Impact on EL – Lending

In this part of the analysis, we are going to examine the impact of the transmission channels from 1 to 3 as described in Section 3.

5.1.1. Description of the Analysis

The first and most immediate impact of increased accuracy of rating systems on bank profitability is related to the loan origination phase. Using rating models that are more predictive at the origination stage makes it possible to optimise the quality of the loan portfolio, granting fewer loans to counterparties that will go into default and at the same time increasing margins on performing customers.

To quantify the impact of the AUROC on the Profit and Loss of the credit portfolio according to this transmission channel, we assume that, according to the Bank's credit policies, out of the 10,000 loan applications the Bank receives, it is able to finance the best 5,000 loans (i.e. the first 5,000 candidate loans with lowest risk in terms of Predicted PD).

Once the Bank selects the 5,000 lowest-predicted-risk loans (based on calibrated PDs), it calculates the default rate of the selected 5,000 loans. Our objective is to estimate the different default rates observed when the lending decision is based on the predictions from models with different AUROC levels.

---

*[**START**] TECHNICAL DESCRIPTION OF THE QUANTIFICATION PROCEDURE*

Steps followed:

- Start from simulated credit portfolios using logistic score distributions calibrated to produce six distinct AUROC bands: 64-66%, 69-71%, 74-76%, 79-81%, 84-86%, and 89-91%.

- Each simulation included:
  - 9,000 non-defaults and 1,000 defaults (10,000 total observations)
  - A target average PD of 10% (portfolio-level)
  - Calibration via logit-shift to match the target PD

- For each AUROC band, 100 valid simulations were selected (where AUROC was within the target range).
- For each simulation:
    - A DataFrame was created with "calibrated PD predictions"[3] and the true default flag.
    - Observations were sorted in ascending order by PD.
    - The 5,000 observations with the lowest PDs were selected.
    - The number of actual defaults among these 5,000 was recorded.
  - Finally, we computed the average number of defaults and default rate among the "best 5,000" loans across all simulations in that AUROC band.

---

*[**END**] TECHNICAL DESCRIPTION OF THE QUANTIFICATION PROCEDURE*

This methodology allows us to assess how model performance influences credit quality in the top-rated segment of the loan book.

---

[3] *For calibrated to be equal to 3% in all simulation in all AUROC levels.*

5.1.2. Results

Based on the simulation setting described above, we tried to answer the following question: how many defaults is the Bank finance (i.e., how many new defaults will it finance each year) if it bases its lending decisions on models with AUROC around 65, 70, 75, 80, 85, or 90?

The answer is provided in the following table:

Table 5.1: Default Mean and Rate per AUROC Range

AUROC Range	Default Count (Mean[4])	Default Rate (%)
64 - 66%	296.82	5.95%
69 - 71%	233.49	4.65%
74 - 76%	182.41	3.64%
79 - 81%	128.63	2.60%
84 - 86%	78.86	1.61%
89 - 91%	42.04	0.82%

As model AUROC improves, the default rate among the best-scoring 5,000 loans decreases substantially. This highlights the importance of model discriminatory power in identifying truly low-risk borrowers. For example, at an AUROC of 64-66%, nearly 6% of the best-scoring loans still default, whereas at 89-91%, the default rate drops to below 1%.

This confirms the common intuition that a better AUROC results in significantly more effective risk assessments, leading to lower expected losses from the top-rated (most creditworthy) borrowers, directly impacting lending profitability.

---

[4] *Mean over 100 simulations made for each AUROC level.*

5.2. Impact on EL – Adverse Selection Effect

5.2.1. Description of the Analysis

The previously quantified effect is only partial, because what happens in the real world also depends also on customer behaviour.

In fact, customers looking for a loan will approach more than one Bank for the same loan and will choose only the one that offers them the best conditions.

In other words, even if a bank decides to finance customer X (performing), that customer may not confirm the request if they receive approval from another Bank under better conditions. Generally, loan conditions depend on the predicted risk of the loan. Therefore, once again, the predictive power of the models plays a crucial role in attracting better (performing) customers with low risk estimates, and discouraging worse (defaulting) customers with higher risk estimates (or at least higher than those predicted by competitors). If even some of the good clients change banks, the return of our Bank's portfolio will be correspondingly lower.

We tried to quantify this effect by introducing a "Customer Probability to Leave" which is a function of the difference between the rating class assigned to the customer by our Bank and the lowest rating class assigned by other Banks in the market:

- If this difference is zero, we assume the customer will accept the loan offer.
- Otherwise, the customer is assumed to decline the offer with a probability that increases as the difference between our Bank's rating class and the lowest one offered in the market increases.

In practice, this approach is more realistic, but also more sensitive to the assumptions about customer behaviour; therefore, care must be taken to define and validate the weighting logic (e.g., how the difference in the rating classes maps to customer churn risk) to avoid biased or misleading estimates of portfolio quality.

To the best of our knowledge, no research has been published on the empirical relation between credit spreads (which we assume to be directly related to the difference between rating classes assigned by the different models) and customer behaviour. Therefore, we tested the following distributions of the Probability to Stay:

Table 5.2: Client Retention by Rating Class under Adverse Selection Scenarios

Rating Class	All Stay	All Leave	Linear Decrease	Exponential Decrease	S-Curve Decrease	High Friction Market
0	1.0	0.0	1.0	1.0	1.0	1.0
1	1.0	0.0	0.8	0.7	0.95	0.95
2	1.0	0.0	0.6	0.5	0.85	0.9
3	1.0	0.0	0.4	0.3	0.6	0.85
4	1.0	0.0	0.2	0.15	0.3	0.8
5	1.0	0.0	0.1	0.1	0.1	0.75
6	1.0	0.0	0.1	0.05	0.05	0.7
7	1.0	0.0	0.1	0.03	0.03	0.65
8	1.0	0.0	0.1	0.02	0.02	0.6
9	1.0	0.0	0.1	0.01	0.01	0.55
10	1.0	0.0	0.1	0.01	0.01	0.5

The above table can be interpreted as follows:

- Under the "**All Stay**" hypothesis, there is no Adverse Selection effect, meaning that regardless of the difference between the Rating Class assigned by a Bank and the minimum Rating Class a client found in the market, then that client will not leave that Bank. In other words, the Probability to Stay with Bank (i.e. 1 - Probability to Leave) is always 100%;
- Under the "**All Leave**" hypothesis, the Adverse Selection effect is extreme, and all the Clients will leave the Bank in any case (this is done only for technical simulation). In other words, the Probability to Stay with Bank (i.e. 1 - Probability to Leave) is always 0%;
- Between these two extreme cases, we tested several alternative hypotheses, identifying the "**S-Curve Decrease**" hypothesis as the most plausible. "S-Curve Decrease" hypothesis provides very low Probabilities to Leave in case of small differences between the Rating Class assigned by a Bank and the

minimum Rating Class a client found in the market (due to, for example, transaction costs the client may face in the real market when moving from one Bank to another). These Probabilities to Leave increase significantly when the difference is greater than or equal to 3 rating classes, such that, in case of differences greater than or equal to 7 rating classes, the Probability to Stay is below 5%, meaning that it is very likely that the client will leave the Bank for another Bank which has offered significantly better conditions.

To assess the expected effect of adverse selection on the credit portfolio, each of the 10,000 Prospect Customers and their possible default flag was weighted by their Probability to Stay (i.e., 1 - the Probability to Leave). Therefore, a customer X with a 20% Probability to Leave will no longer be weighted as 1, as in the previous analysis, but rather as 0.8, since there is a 20% chance they will no longer be a client of a Bank and will turn to another institution. Moreover, if customer X is a default, the weight of that default will also be 0.8 (instead of 1) for the purpose of calculating the Default Rate of new loans.

Then, consistent with the earlier assumption that the Bank can issue 5,000 loans based on its annual Credit Policies, we calculated the new Default Rate after incorporating the adverse selection effect by sorting customers in ascending order of predicted PD and identifying the new portfolio as the one in which the sum of the weighted customer weights equals 5,000.

---

*[**START**] TECHNICAL DESCRIPTION OF THE QUANTIFICATION PROCEDURE*

Steps followed:

- For each AUROC band (64-66% to 89-91%), we analysed 100 simulations of calibrated score distributions (each containing 10,000 observations with a 10% default rate).
- Each observation has:
    - A calibrated PD, and
    - An associated *adjusted loan weight* (Weight_adj by the Probability to Leave of the specific loan)*

    - For each simulation:
        - The dataset was sorted by ascending PD, simulating a lender approving the lowest-risk borrowers first.
        - Loans were accumulated until the sum of adjusted weights reached ~5000 units.
        - The resulting "accepted portfolio" was analysed to compute:
            - Total defaults (Flag_def sum)
            - Default rate (defaults over total population, i.e., 10,000)
            - Number of accepted loans
- Final metrics (averages across 100 simulations per AUROC group) were calculated to quantify the impact of model performance on EL under fixed-lending-volume constraints.

As mentioned above, the *"adjusted loan weight (Weight_adj)"* quantifies the stability of a loan recommendation for a given client, indicating both the lender's exposure and the likelihood that the client will continue to qualify for the loan under the current risk assessment. It is derived through the following steps:

1. **Probability of Default (PD) Calibration**: Each client's risk is first expressed as a calibrated Probability of Default (PD_calibrated), estimating their chance of repayment failure.
2. **Risk Rating Assignment**: This PD is mapped to an internal risk rating scale (e.g., AAA, BB, etc.), similar to credit ratings, to categorise clients by risk level.
3. **Lowest Acceptable Rating Determination**: For each model performance level (measured by AUROC, which evaluates risk discrimination), the system identifies the minimum allowable rating that still qualifies for the loan.
4. **Rating Comparison and Gap Analysis**: The client's assigned rating is compared to this minimum threshold. The difference between the two reflects how much safer (or riskier) the client is relative to the cutoff.
5. **Probability Mapping**: This rating gap is converted into a probability (0–100%) representing the client's likelihood of retaining the loan recommendation. A smaller gap (e.g., a client near the threshold) results in a lower probability, indicating higher instability.

6. **Weight_adj Output**: The final probability becomes Weight_adj, serving as a stability metric:
    a. High Weight_adj (>70%): The loan recommendation is reliable and unlikely to change.
    b. Low Weight_adj (<30%): The recommendation is borderline and may be revised if risk thresholds shift.

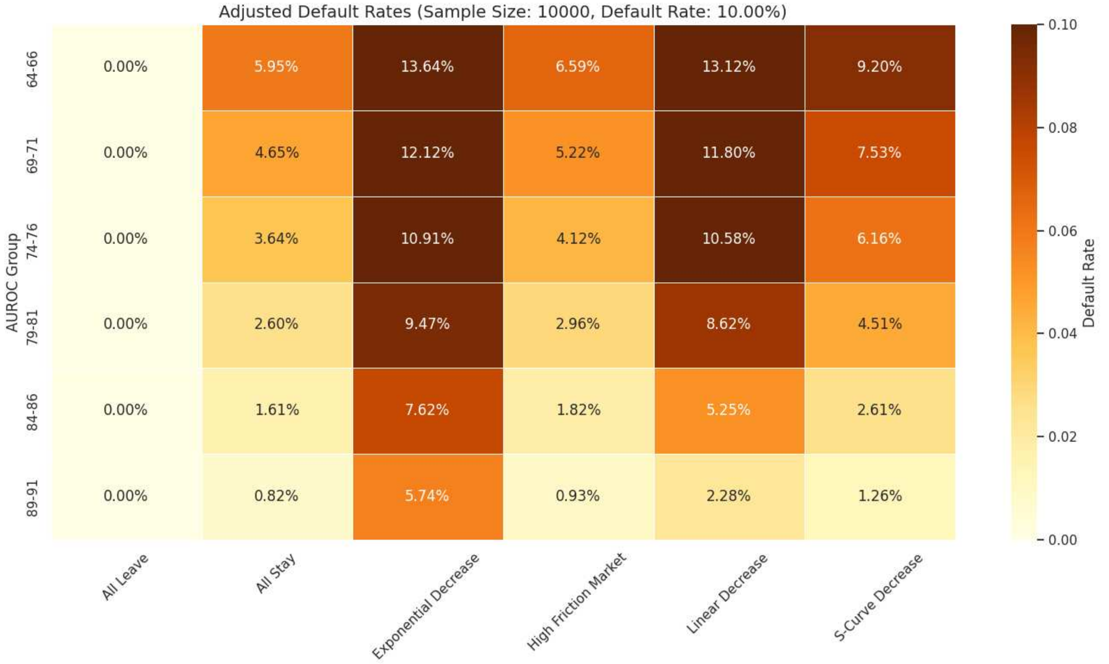
---

*[**END**] TECHNICAL DESCRIPTION OF THE QUANTIFICATION PROCEDURE*

5.2.2. Results

The resulting default rates under each scenario (Probability to Leave distribution) have been summarised in the following picture:

Figure 5.1: Adjusted Default Rates (Sample Size: 10000, Default Rate: 10.00%)



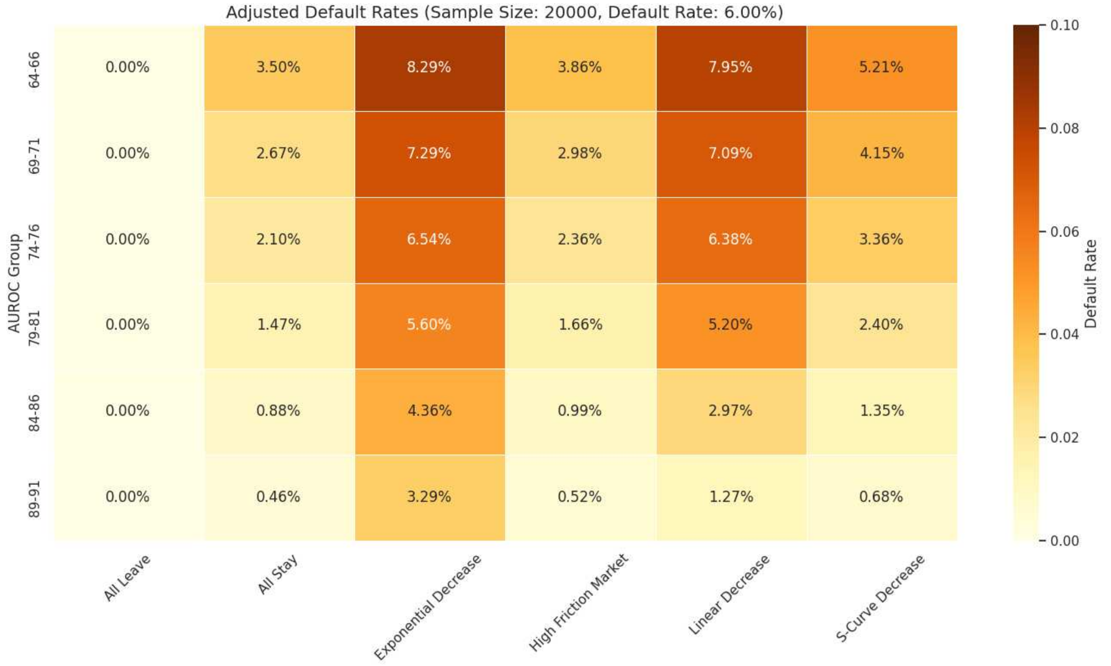
AUROC Group	All Leave	All Stay	Exponential Decrease	High Friction Market	Linear Decrease	S-Curve Decrease
64-66	0.00%	5.95%	13.64%	6.59%	13.12%	9.20%
69-71	0.00%	4.65%	12.12%	5.22%	11.80%	7.53%
74-76	0.00%	3.64%	10.91%	4.12%	10.58%	6.16%
79-81	0.00%	2.60%	9.47%	2.96%	8.62%	4.51%
84-86	0.00%	1.61%	7.62%	1.82%	5.25%	2.61%
89-91	0.00%	0.82%	5.74%	0.93%	2.28%	1.26%

The heatmap illustrates adjusted default rates across various AUROC performance groups (from 64-66 to 89-91) under different behavioural assumptions or interventions, using a sample size of 10,000 and a baseline (market) default rate of 10%.

As expected, default rates decline with higher AUROC values, demonstrating the effectiveness of better classification models in isolating risk. Notably, the "Exponential Decrease" and "Linear Decrease" scenarios show the highest sensitivity to model improvement, with default rates dropping steeply as AUROC increases - indicating their strong reliance on predictive accuracy. In contrast, the "All Leave" group maintains a 0% default across all AUROCs, reflecting the removal of the entire risky population. Meanwhile, the "All Stay" and "High Friction Market" scenarios show moderate reductions in default rates, suggesting limited benefit from predictive gains under those retention strategies.

Below, another parametrisation is depicted for the same analysis, by changing both the initial sample size of the simulations from 10,000 to 20,000 and the market Default Rate from 10% to 6%, but still keeping the number of simulations to 100.
It is important to note that we consider this parametrization with a 6% market default rate to be rather unrealistic, except in periods of strong economic expansion. Nonetheless, simulation is useful to assess the minimum potential benefit that can be obtained from adopting more accurate models.

Figure 5.2: Adjusted Default Rates (Sample Size: 20000, Default Rate: 6.00%)



As with the previous figure, improved AUROC scores correspond to consistently lower default rates, confirming that better model discrimination effectively concentrates risk.

The “Exponential Decrease” and “Linear Decrease” strategies again exhibit the steepest gradient of improvement, dropping from 8.29% to 3.29% and from 7.95% to 1.27%, respectively, as AUROC improves - highlighting the benefits of strategic retention paired with predictive strength. “All Leave” continues to show a flat 0.00% default, while “All Stay” maintains moderate levels regardless of AUROC, indicating limited responsiveness to model improvements.

Compared to the previous plot (with 10% baseline defaults), the absolute default rates are lower here, but the relative patterns persist as expected.

Stronger predictive models (higher AUROC) **significantly reduce Default Rates** for a fixed lending budget by avoiding adverse selection. As model AUROC increases, lenders can **select fewer but higher-quality borrowers**, concentrating exposure in safer segments and **reducing default rates by up to 85%** between the worst and best model groups.

5.3. Impact on RWA

The third transmission channel identified concerns RWA savings attributable to the greater discriminatory power of rating models. Here, we analysed two scenarios:

(1) a Bank that already has a rating scale in place and does not intend to modify it due to internal constraints (e.g., IT or business-related). This represents the standard and most frequent case. Nevertheless, to the best of our knowledge, it has not been analysed in previous studies;

(2) a Bank that has the opportunity to define its rating scale from scratch. This case is rarer and applies only to traditional Banks transitioning from a Standardized approach to an IRB approach for capital requirement calculation, or to new Banks that choose to adopt an IRB approach from the outset for capital requirement purposes.

5.3.1. RWA Savings in case of rating systems already in place

In the first part of the analysis, we are going to calculate the RWA given that the financial institution has already applied an IRB rating system.

This analysis explores the relationship between the discriminatory power of a credit risk model (measured by AUROC) and its impact on Risk-Weighted Assets (RWA). In regulatory capital frameworks such as Basel II/III, Banks can use internal rating systems to estimate capital requirements. A model with a higher AUROC better separates defaulters from non-defaulters, which theoretically leads to more accurate Probability of Default (PD) estimates. This, in turn, can lead to lower and more risk-sensitive RWA values. Understanding how incremental improvements in model performance (via AUROC) translate into capital savings (via RWA reductions) is key for risk model validation, optimisation, and regulatory approval.

As described in section 3, the savings in RWA that can be attributed to higher rating models' AUROC do not derive from any mathematical constraint; they are just "expected" as a result of two main factors:

1. The first is the shape of the capital requirement function for PD, which is concave - meaning it requires low amounts of capital for both very low and very high PD values. In fact, the capital requirement is zero when PD = 0% as well as when PD = 100%.
2. The second factor is the expected behaviour of models as their discriminatory power increases. To get an intuition for this, one can imagine the behavior of a perfect or ideal model: such a model, if it existed, would always assign a PD = 0% to all clients who will not default in the next 12 months (default flag = 0), and a PD = 100% to all clients who will default (default flag = 1). As a result, the RWA capital requirement would be zero. Models with increasing discriminatory power (i.e., models approaching perfection), being increasingly capable of distinguishing between good and bad clients, will tend to concentrate their predictions at the extremes of the distribution. This, in turn,

leads to a reduction in the required RWA capital. This effect is present even when the Bank uses the same rating scale.

The objective of our analysis is to quantify the savings in RWA that a Bank could expect by increasing the accuracy of its rating models.

5.3.2. Description of the Analysis

The process can be described step by step as follows:

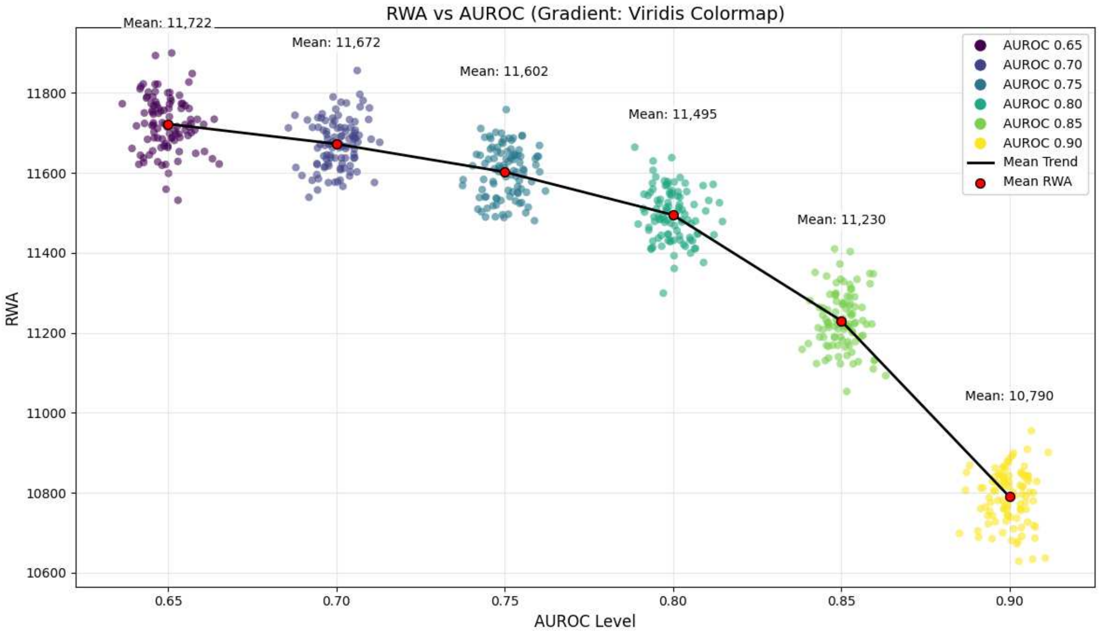
1. **Calculate the RWA[5] requirement by AUROC levels**
    - RWA results are computed by AUROC level given some fixed parameters (LGD = 15% and EAD = 1) according to the rating scale defined for our simulation setting (thresholds are taken from credit rating breakpoints used for EU Banks Pillar 3 reporting, e.g. <0.0015, 0.0015-0.0025, ..., >0.3).
    - RWA results are collected for our range of AUROC values (e.g., 0.65, 0.70, ..., 0.90).
    - Each group represents the RWAs corresponding to models with a specific AUROC.
2. **Plot individual RWA points.**
    - Each RWA value is plotted on the vertical axis, with AUROC on the horizontal axis.
3. **Calculate the mean RWA for each AUROC group**
    - For every AUROC level, the average (mean) RWA is computed.
4. **Plot the mean RWA trend.**
    - Mean points are marked with red circles, and a black line connects them to show the trend. This helps summarise the relationship between AUROC and average RWA.

---

[5] *We used the Retail Mortgages RWA formula which we believe is consistent with the assumptions made throughout this paper and, in particular, the availability of a granular (retail) portfolio with low risk level. In case of higher risk portfolios, RWA results, and savings are expected to be also higher.*

5.3.3. Results

Figure 5.3: RWA Evolution with 5-Percentile Increases in AUROC



The graph shows a clear and smooth negative relationship between AUROC and RWA: as the AUROC increases from 0.65 to 0.90, the average RWA decreases steadily from around 11,722 to 10,790.

This supports the hypothesis that better-performing models (higher AUROC) can lead to lower capital requirements, reflecting more accurate and risk-sensitive credit assessments. The decrease, however, is not linear; the slope steepens at higher AUROC levels, suggesting that improvements in model quality beyond a certain threshold can result in increasingly larger capital benefits. This has strong implications for Banks considering investment in improving their risk models, since even marginal increases in AUROC beyond 0.80 may yield meaningful RWA reductions.

5.4. RWA Savings in case of newly developed rating systems

In this scenario, the financial institution is also interested in creating a new rating system. This usually happens for new or small institutions that are usually more flexible to build their own rating systems based on their portfolios.

5.4.1. Description of the Analysis

This study examines how the predictive accuracy of credit risk models - measured by the AUROC - influences the calculation of RWA under the Basel regulatory framework. The goal is to assess whether better-performing models lead to more efficient capital allocation by reducing unnecessary overestimation of risk.

The analysis begins by processing simulated datasets, each representing different levels of model performance (AUROC ranges from 64–66 up to 89–91). For each dataset, the predicted probabilities of default (PDs) are sorted and divided into 50 initial buckets of equal size. However, regulatory standards require that these buckets meet two key conditions:

1. **The 2% Rule**: The absolute difference between the average predicted PD and the actual observed default rate in each bucket should not exceed 2%. If a bucket violates this rule, it is merged with adjacent buckets until compliance is achieved.
2. **Monotonicity Enforcement**: Default rates must increase as PDs increase. If a bucket has a lower observed default rate than the one before it, they are merged to ensure a logical risk progression.

Once the final buckets are established, the Basel capital requirement formula is applied to compute RWA. This formula considers the probability of default, loss given default (fixed at 20%), and a correlation factor (set at 0.15), adjusted for a 99.9% confidence level. The resulting capital requirement is then scaled by a conversion factor (12.5x) to determine the total RWA contribution for each bucket.

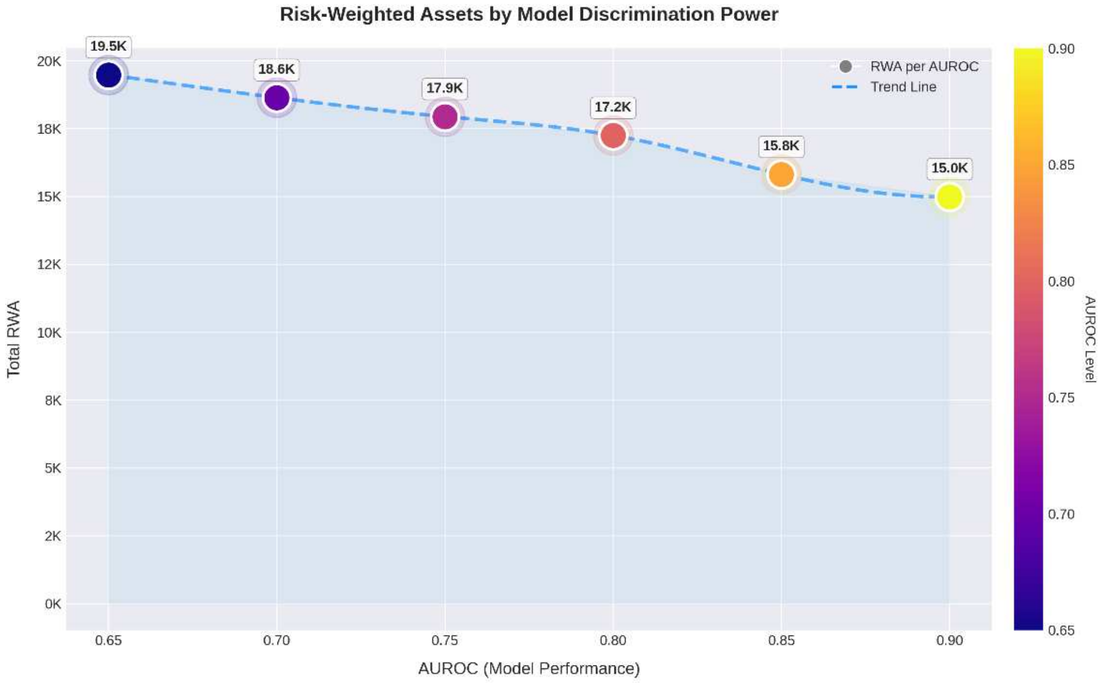
To visualise the results, two key plots are generated for each AUROC scenario:

- **PD Calibration Plot**: Compares the mean predicted PD in each bucket against the actual observed default rate, illustrating how well the model's predictions align with reality.
- **RWA Distribution Plot**: Shows the capital allocation across risk buckets, highlighting which segments contribute most to the total RWA.

By comparing these results across different AUROC levels, we can assess whether more accurate models lead to better-calibrated risk estimates and, consequently, more precise capital requirements.

5.4.2. Results

Figure 5.4: RWA vs AUROC for a single simulation



- **PD Calibration Accuracy**: Higher AUROC models (e.g., 89-91) demonstrate a stronger alignment between predicted PDs and observed default rates, with points clustering closely around the ideal 45-degree line. In contrast, lower AUROC models (e.g., 64-66) show more dispersion, often requiring aggressive bucket merging to comply with the 2% rule.
- **RWA Allocation**: The bulk of RWA contributions typically comes from high-PD buckets, as expected, but the distribution varies with model performance. Better models tend to assign capital more discriminately, avoiding excessive RWA in moderate-risk segments due to overestimation.
- **Total RWA Impact**: While the exact relationship depends on the dataset, there is a general trend where higher AUROC models lead to slightly lower total RWA, as they reduce the mispricing of risk. However, the magnitude of this effect depends on the underlying risk distribution in the portfolio.

For example, one simulation might yield:

"Total RWA for AUROC 64-66: 19,500€"
"Total RWA for AUROC 89-91: 15,000€"

This suggests that improved model discrimination can lead to capital savings, though the extent varies. The bucket-level analysis further highlights which risk segments drive these differences, which is a critical piece of information for risk managers optimising capital allocation.

Ultimately, the analysis demonstrates that model performance directly affects regulatory capital efficiency. However, it also underscores that even well-performing models must undergo rigorous calibration adjustments to meet Basel requirements.

CHAPTER 6

PROFITABILITY METRICS

In this section, we attempt to translate the technical metrics of AUROC and Default Rate into meaningful economic values, namely the Bank's additional profit and rate of return.

To do so, we adopted the following baseline parametrization, assuming a small, low-risk retail portfolio (e.g., the mortgage portfolio of a small Bank).

Table 6.1: Baseline Parameters for Economic Impact Estimation

PARAMETER	VALUE
NUMBER OF LOANS (size of the current Bank portfolio)	50.000
STARTING DEFAULT RATE (current portfolio)	3%
AVG EXPOSURE PER LOAN (Euro)	70.000
TOTAL EXPOSURE (Euro)	3.500.000.000
NEW LOANS GRANTED PER YEAR	5.000
LGD	15%
BANK MARGIN ON PERFORMING EXPOSURES	1%
NUMBER OF PROSPECT CLIENTS (market size)	10.000
DEFAULT RATE OF PROSPECT CLIENTS (market risk)	10%

The following tables report the effect of improved rating model accuracy on economic fundamentals, resulting from the parameterization described above and the default rates described in *section 5.2.2*.

The first profitability metric we calculated is the “Delta profit”, defined as the simple difference in profit at time T+1 that a Bank obtains when granting loans based on a credit rating model with AUROC in range X (e.g., 70), compared to the profit at time T+1 obtained using a model with AUROC in the immediately lower range X-1 (65 to stick with the previous example). Results are differentiated according to the mapping used.

Table 6.2: Annual Delta Profit by AUROC range and Adverse Selection Scenario

	Delta Profit	Delta Profit	Delta Profit	Delta Profit	Delta Profit
AUROC	**All stay**	**Exponential Decrease**	**High Friction Market**	**Linear Decrease**	**S-Curve Decrease**
65-70	728.000	851.200	767.200	739.200	935.200
70-75	565.600	677.600	616.000	683.200	767.200
75-80	582.400	806.400	649.600	1.097.600	924.000
80-85	554.400	1.036.000	638.400	1.887.200	1.064.000
85-90	442.400	1.052.800	498.400	1.663.200	756.000

The results vary depending on the type of mapping used to model the adverse selection effect observed in real-world markets.

We consider the most realistic simulation to be the "S-curve decrease" (highlighted in blue in the table above), in which small differences between the rating classes assigned by the Bank and those assigned by the market result in very low probabilities of customer churn (e.g., due to transaction costs - not necessarily monetary - that offset the economic benefit for the customer) but, when the rating differences become larger (beyond three classes), the probability of leaving the Bank increases significantly.

In this "S-curve decrease" simulation, after just one year, if a Bank originates new loans (which, under our conservative baseline parametrization, represent only 10% of the total portfolio - 5,000 new loans in a 50,000-loan portfolio) using a model with 5 AUROC points higher predictive accuracy, it achieves an increase in profits of approximately €800,000 to €1 million (on a €3.5 billion portfolio, that is a small size portfolio) per year.

It is important to note that these effects accumulate year after year, assuming that the performance gap between models remains constant.

The relative impact of higher model accuracy becomes clearer when computing the percentage increase in profit, for example, comparing “Profit at T+1 using a model with AUROC 70” vs “Profit at T+1 using a model with AUROC 65” (this is a % difference, i.e. it is computed as “Profit from AUROC 70” - “Profit from AUROC 65” / “Profit from AUROC 65”).
The profit increases range from a minimum of 2.3% when moving from an AUROC of 85 to 90 under the assumption of no adverse selection effect (“all stay”), up to +12.5% after just one year when improving AUROC from 80 to 85 and accounting for adverse selection through a linear decrease function.

Table 6.3: %Profit Increase by AUROC range and Adverse Selection Scenario

	% Profit Increase	% Profit Increase	% Profit Increase	% Profit Increase	% Profit Increase
AUROC	**All stay**	**Exponential Decrease**	**High Friction Market**	**Linear Decrease**	**S-Curve Decrease**
65-70	4,4%	7,0%	4,7%	5,9%	6,3%
70-75	3,3%	5,2%	3,6%	5,1%	4,9%
75-80	3,3%	5,9%	3,7%	7,9%	5,6%
80-85	3,0%	7,1%	3,5%	12,5%	6,1%
85-90	2,3%	6,7%	2,6%	9,8%	4,1%

Finally, if we compute the Bank’s rate of return as Profit divided by Total Exposure (€3,5 BN), we can derive the delta RoR as the simple difference between:

- the RoR at time T+1 achieved by the Bank granting new loans using a model with AUROC “X”, and
- the RoR at time T+1 achieved by the same Bank granting new loans using a model with AUROC “X minus 5 pp”.

Table 6.4: ΔRoR by AUROC Range and Adverse Selection Scenario

AUROC	Delta RoR **All stay**	Delta RoR **Exponential Decrease**	Delta RoR **High Friction Market**	Delta RoR **Linear Decrease**	Delta RoR **S-Curve Decrease**
65-70	0,02%	0,02%	0,02%	0,02%	0,03%
70-75	0,02%	0,02%	0,02%	0,02%	0,02%
75-80	0,02%	0,02%	0,02%	0,03%	0,03%
80-85	0,02%	0,03%	0,02%	0,05%	0,03%
85-90	0,01%	0,03%	0,01%	0,05%	0,02%

The above results do not change significantly if we change the number of prospective clients from 10,000 to 20,000, everything else equal.

As a next step, to obtain an even more prudential estimate of the economic benefits of granting new loans with more accurate rating models, we also tested a slightly different parametrization, where the Default Rate of the prospective clients is 6% instead of 10%.

Naturally, lowering the default rate of the market from which the Bank selects its customers also reduces the benefit of using more accurate models (in the extreme case, if the market default rate were 0%, there would be no advantage in employing higher-performing models).

It is important to note that we consider this parametrization with a 6% market default rate to be rather unrealistic, except in periods of strong economic expansion. Nonetheless, simulation is useful to assess the minimum potential benefit that can be obtained from adopting more accurate models.

The following alternative parametrization has been adopted.

Table 6.5: Alternative Portfolio Parametrization (DR of Prospect Clients)

PARAMETER	VALUE
NUMBER OF LOANS (size of the current Bank portfolio)	50.000
STARTING DEFAULT RATE (current portfolio)	3%
AVG EXPOSURE PER LOAN (Euro)	70.000
TOTAL EXPOSURE (Euro)	3.500.000.000
NEW LOANS GRANTED PER YEAR	5.000
LGD	15%
BANK MARGIN ON PERFORMING EXPOSURES	1%
NUMBER OF PROSPECT CLIENTS (market size)	20.000
DEFAULT RATE OF PROSPECT CLIENTS (market risk)	**6%**

The following tables report the effect of improved rating model accuracy on economic fundamentals, resulting from the parameterization described above and the default rates described in *section 5.2.2*.

Table 6.6: Delta Profit Impact by AUROC and Market Assumptions

	Delta Profit	Delta Profit	Delta Profit	Delta Profit	Delta Profit
AUROC	**All stay**	**Exponential Decrease**	**High Friction Market**	**Linear Decrease**	**S-Curve Decrease**
65-70	464.800	560.000	492.800	481.600	593.600
70-75	319.200	420.000	347.200	397.600	442.400
75-80	352.800	526.400	392.000	660.800	537.600
80-85	330.400	694.400	375.200	1.248.800	588.000
85-90	235.200	599.200	263.200	952.000	263.200

Table 6.7: %Profit Increase by AUROC and Market Assumptions

	% Profit Increase	% Profit Increase	% Profit Increase	% Profit Increase	% Profit Increase
AUROC	**All stay**	**Exponential Decrease**	**High Friction Market**	**Linear Decrease**	**S-Curve Decrease**
65-70	2,6%	3,7%	2,8%	3,1%	3,5%
70-75	1,7%	2,7%	1,9%	2,5%	2,5%
75-80	1,9%	3,2%	2,1%	4,1%	3,0%
80-85	1,7%	4,1%	2,0%	7,4%	3,2%
85-90	1,2%	3,4%	1,4%	5,2%	1,4%

Table 6.8: Delta RoR Impact by AUROC and Market Assumptions

	Delta RoR	Delta RoR	Delta RoR	Delta RoR	Delta RoR
AUROC	**All stay**	**Exponential Decrease**	**High Friction Market**	**Linear Decrease**	**S-Curve Decrease**
65-70	0,01%	0,02%	0,01%	0,01%	0,02%
70-75	0,01%	0,01%	0,01%	0,01%	0,01%
75-80	0,01%	0,02%	0,01%	0,02%	0,02%
80-85	0,01%	0,02%	0,01%	0,04%	0,02%
85-90	0,01%	0,02%	0,01%	0,03%	0,01%

As shown in the previous tables, even under a conservative parametrization of the risk of the prospective clients, the effect of increased accuracy is still significant.

Every year (since such benefits can be accumulated annually), the Bank's profit appears to increase by approximately €0.5 million on average for each performance achieved (i.e. for each 5% increase in AUROC) when computed on a relatively small portfolio of €3.5 billion. Also, the *% Profit Increase* is quite high, ranging from an average 1.8% increase without considering any adverse selection effect (“all stay” simulation) to an average 4.5% increase under the “Linear Decrease” simulation.

We have noticed that our results are nonetheless lower than those reported by Jankowitsch et al., who observe an increase of 30-40 basis points in the Rate of Return when moving from a low-accuracy model to a medium-accuracy model, and an additional increase of 15 basis points when moving from a medium-accuracy model to a high-accuracy model. Although these results are not directly comparable with ours due to differences in the estimation approach (e.g. our model performance is measured in terms of AUROC, whereas Jankowitsch et al. use the error rate added to the counterparties' true PDs), we attempted to align our baseline parametrization with that of Jankowitsch et al.

To do so, we assumed that the Jankowitsch et al. transition from a “low-performance model to a medium-performance model” and from a “medium-performance model to a high-performance model” can be interpreted, in our framework, as improvements of 15 AUROC points. In other words, we assumed that moving from a low to a medium performance model corresponds to an increase in AUROC from 65 to 80, while the transition from medium to high performance corresponds to an increase from 75 to 90.

We therefore modified the baseline parametrization as follows, aligning it with the values reported in the paper by Jankowitsch et al. (in blue, the differences with respect to our base parametrization).

Table 6.9: Conservative Parameterization (LGD, MARGIN)

PARAMETER	VALUE
NUMBER OF LOANS	50.000
STARTING DEFAULT RATE (current portfolio)	3%
AVG EXPOSURE PER LOAN (Euro)	70.000
TOTAL EXPOSURE (Euro)	3.500.000.000
NEW LOANS GRANTED PER YEAR	5.000
LGD	45%
MARGIN ON PERFORMING EXPOSURES	5%
NUMBER OF PROSPECT CLIENTS	10.000
DEFAULT RATE OF PROSPECT CLIENTS	10%

Based on this less conservative parametrization, we obtain the following delta RoRs (which is the profitability metric reported by Jankowitsch et al.).

Table 6.10: Delta RoR by AUROC Range and Adverse Selection Scenario

	Delta RoR	Delta RoR	Delta RoR	Delta RoR	Delta RoR
AUROC	**All stay**	**Exponential Decrease**	**High Friction Market**	**Linear Decrease**	**S-Curve Decrease**
65-70	0,06%	0,07%	0,07%	0,06%	0,08%
70-75	0,05%	0,06%	0,05%	0,06%	0,07%
75-80	0,05%	0,07%	0,06%	0,09%	0,08%
80-85	0,05%	0,09%	0,05%	0,16%	0,09%
85-90	0,04%	0,09%	0,04%	0,14%	0,06%

Under the most realistic mapping hypothesis ("S-Curve Decrease"), we obtain an increase of 23 basis points moving from a low-performance (AUROC=65) to a medium-performance (AUROC=80) model, and another increase of 23 basis points moving from a medium-performance (AUROC=75) to a high-performance (AUROC=90) model.

These results, although not perfectly matching those reported by Jankowitsch et al., are reasonably consistent, indicating a Delta RoR of approximately 50 basis points when moving from a low-performance to a high-performance model in both simulations.

Finally, we incorporated the benefits of improved loan selection by also including those derived from the impact of AUROC on the Bank's Risk-Weighted Assets (RWA), as from *section 5.3.3*. For this simulation, we assumed a Bank cost of capital equal to 5%[6].

---

[6] *Source https://pages.stern.nyu.edu/~adamodar/New_Home_Page/datafile/wacc.html*

Table 6.11: Impact of AUROC Improvement on RWA and Cost of Capital Savings

AUROC LEVEL	AVG RWA	CAPITAL (RWA * 8%)	COST OF CAPITAL (CAPITAL * 5%)	COST SAVINGS
65	11.722.000.000	937.760.000	46.888.000	-
70	11.672.000.000	933.760.000	46.688.000	200.000
75	11.602.000.000	928.160.000	46.408.000	280.000
80	11.495.000.000	919.600.000	45.980.000	428.000
85	11.230.000.000	898.400.000	44.920.000	1.060.000
90	10.790.000.000	863.200.000	43.160.000	1.760.000

If we include these additional savings coming from RWA reduction into our simulated Bank P&L, we obtain the following profitability metrics.

Table 6.12: Delta Profit Impact by AUROC for Conservative Parametrization

	Delta Profit	Delta Profit	Delta Profit	Delta Profit	Delta Profit
AUROC	**All stay**	**Exponential Decrease**	**High Friction Market**	**Linear Decrease**	**S-Curve Decrease**
65-70	928.000	1.051.200	967.200	939.200	1.135.200
70-75	845.600	957.600	896.000	963.200	1.047.200
75-80	1.010.400	1.234.400	1.077.600	1.525.600	1.352.000
80-85	1.614.400	2.096.000	1.698.400	2.947.200	2.124.000
85-90	2.202.400	2.812.800	2.258.400	3.423.200	2.516.000

Table 6.13: %Profit Increase by AUROC for Conservative Parametrization

	% Profit Increase	% Profit Increase	% Profit Increase	% Profit Increase	% Profit Increase
AUROC	**All stay**	**Exponential Decrease**	**High Friction Market**	**Linear Decrease**	**S-Curve Decrease**
65-70	5,6%	8,6%	6,0%	7,5%	7,7%
70-75	4,9%	7,3%	5,3%	7,3%	6,7%
75-80	5,7%	9,0%	6,1%	10,9%	8,2%
80-85	8,8%	14,4%	9,3%	19,6%	12,2%
85-90	11,6%	18,0%	12,0%	20,2%	13,7%

Table 6.14: Delta RoR by AUROC for Conservative Parametrization

	Delta RoR	Delta RoR	Delta RoR	Delta RoR	Delta RoR
AUROC	**All stay**	**Exponential Decrease**	**High Friction Market**	**Linear Decrease**	**S-Curve Decrease**
65-70	0,03%	0,03%	0,03%	0,03%	0,03%
70-75	0,02%	0,03%	0,03%	0,03%	0,03%
75-80	0,03%	0,04%	0,03%	0,04%	0,04%
80-85	0,05%	0,06%	0,05%	0,08%	0,06%
85-90	0,06%	0,08%	0,06%	0,10%	0,07%

Our analyses demonstrate how enhancing the predictive accuracy of credit risk models creates tangible economic benefits for banks through three core transmission channels. Under the “S-Curve decrease” simulation, the economic benefit of a 1 percentage point increase in AUROC translates, on average, into €330,000 in additional profits each year on a €3.5 billion portfolio. In relative terms, this corresponds to an average 1.9% increase in the Bank's profits.

On the other hand, it is important to note that the development and maintenance of credit rating models entail costs that vary significantly depending on the Bank and the specific characteristics of each portfolio for which the models are built.
These costs are generally well known to Banks, so we did not consider it necessary to include a dedicated simulation for them. Nevertheless, it is easy to observe that even in the presence of substantial development costs (around €1 million), and under the "All stay" simulation scenario (i.e., the one that does not account for any Adverse Selection effect in the market which would further amplify the benefits of higher accuracy), the Return on Investment (from now on ROI) of an investment in a model that yields an improvement of 5 percentage points in AUROC exceeds 100%, even when applied to a small and low-risk portfolio as in our simulation (€3.5 billion).

Furthermore, it is important to reiterate that our estimation represents a prudential quantification, as it focuses only on the most measurable impacts (lending quality, adverse selection, and RWA efficiency) while excluding other important but harder-to-quantify effects mentioned in Section 3, such as monitoring benefits or regulatory add-on reductions.
Moreover, our estimates can be considered conservative due to the parametrization adopted, which reflects a relatively small portfolio (€3.5 billion - comparable to the size of the mortgage portfolio of a small bank. For reference, the retail mortgage portfolios of medium-sized Italian banks such as MPS or BBPM are typically around €25-30 billion, with quite long duration (only 10% of the portfolio is renewed each year, meaning that the effect of the higher accuracy is computed on 10% of the loans) and low-risk characteristics. In the presence of higher default rates and shorter duration, one would expect more accurate models to demonstrate their benefits even more clearly, and the resulting RWA to be generally higher.

We believe that these results represent a strong incentive for Banks to continue investing in the improvement of their credit rating models' accuracy, whether through better estimation techniques or through the use of new data sources.

From another perspective, these effects also capture the economic impact of a potential deterioration in the discriminatory power of rating models, which, if not properly monitored and maintained, can quickly experience performance decay with severely negative consequences on the Bank's profitability.

CHAPTER 7

CONCLUSIONS

This analysis demonstrates how enhancing the predictive accuracy of credit risk models can yield tangible economic benefits for banks through three primary transmission channels: lending quality, adverse selection, and RWA efficiency. The results of the simulations presented in this study offer strong evidence that improvements in model accuracy can generate meaningful ROI, making a compelling case for banks to invest in upgrading their credit risk models.

Under the most realistic simulation scenario, referred to as the "S-Curve Decrease," the economic benefit of a 1 percentage point increase in AUROC translates, on average, into €730,000 in additional profits each year on a €3.5 billion portfolio. In relative terms, this represents an average 1.5% increase in the bank's annual profits. These results are notable given that the analysis only considers the most measurable impacts - namely, improvements in lending quality, the mitigation of adverse selection, and better RWA efficiency. Other potential benefits, such as the reduction in monitoring costs or regulatory capital add-ons, have been excluded due to their more difficult-to-quantify nature, suggesting that the actual economic impact could be even higher than estimated.

Moreover, the conservative assumptions underlying the study, such as a focus on a low-risk portfolio with high duration, further underscore the robustness of the findings. Even in these conservative settings, the impact of improved predictive accuracy is significant, providing a strong financial incentive for banks to invest in better models. The results highlight that even small enhancements in model performance can lead to meaningful improvements in profitability, thus justifying the costs associated with model development and maintenance.

This study also suggests that the economic implications of model accuracy go beyond profitability alone. Enhancements in credit risk models can improve the overall stability of the financial institution, enhance its regulatory standing, and contribute to more efficient capital allocation. Banks that adopt higher-performing models are likely to experience not only direct financial benefits but also an improved risk management framework, helping them navigate the complex and volatile nature of credit markets.

In conclusion, the findings of this analysis provide valuable insights into the cost-benefit trade-offs of credit risk model improvements. They support the notion that investing in advanced credit risk models is not just a technical upgrade but a strategic move that can lead to substantial long-term financial rewards. Furthermore, the study lays the groundwork for future research, which could explore additional channels through which predictive accuracy affects a bank's operations and extend the analysis to larger, more diversified portfolios. As financial institutions continue to face increasing pressures from regulators and the market, enhancing the accuracy of credit risk models will be a critical factor in maintaining competitiveness and ensuring financial sustainability.

CHAPTER 8
BIBLIOGRAPHY

- **Alonso-Robisco & Carbó (2022),** *Can machine learning models save capital for banks? Evidence from a Spanish credit portfolio*, International Review of Financial Analysis

- **Baesens et al. (2015),** *Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research*, European Journal of Operational Research

- **Berg et al. (2018),** *On the Rise of FinTechs: Credit Scoring Using Digital Footprints*, FDIC CFR WP

- **Blöchlinger and Leippold (2005),** *Economic Benefit of Powerful Credit Scoring,* National Centre of Competence in Research Financial Valuation and Risk Management Working Paper 216

- **Brown and Mues (2012)**, *An experimental comparison of classification algorithms for imbalanced credit scoring data sets,* Expert Systems with Applications

- **Fuster et al. (2022),** *Predictably Unequal? The Effects of Machine Learning on Credit Markets,* Journal of Finance

- **Gal et al. (2019),** *How do machine learning and non-traditional data affect credit scoring? New evidence from a Chinese fintech firm,* BIS WP

- **Giudici et al. (2019),** *Network based credit risk models,* Quality Engineering

- **Jankowitsch et al. (2007)**, *Modelling the Economic Value of Credit Rating Systems,* Journal of Banking & Finance

- **Stein (2005),** *The relationship between default prediction and lending profits: Integrating ROC analysis and loan pricing,* Journal of Banking and Finance

- **Zhou and Van Vuuren (2024),** *The impact of deterioration in rating-model discriminatory power on expected losses,* Journal of Risk Model Validation

- **Stephen Satchell and Wei Xia (2006),** *Analytic Models of the ROC Curve: Applications to Credit Rating Model Validation***,** University of Technology Sydney

- **Iftho Hara Khanam (2023),** *Credit scoring using Logistic regression*, Mälardalen University Västerås, Sweden

- **Costa E Silva E, Lopes IC, Correia A, Faria S. (2020),** *A logistic regression model for consumer default risk.*, J Appl Stat.

- **Avramova, Sofiya & Le Lesle, Vanessa (2012).** *IMF Working Paper on Risk-Weighted Assets.* International Monetary Fund.