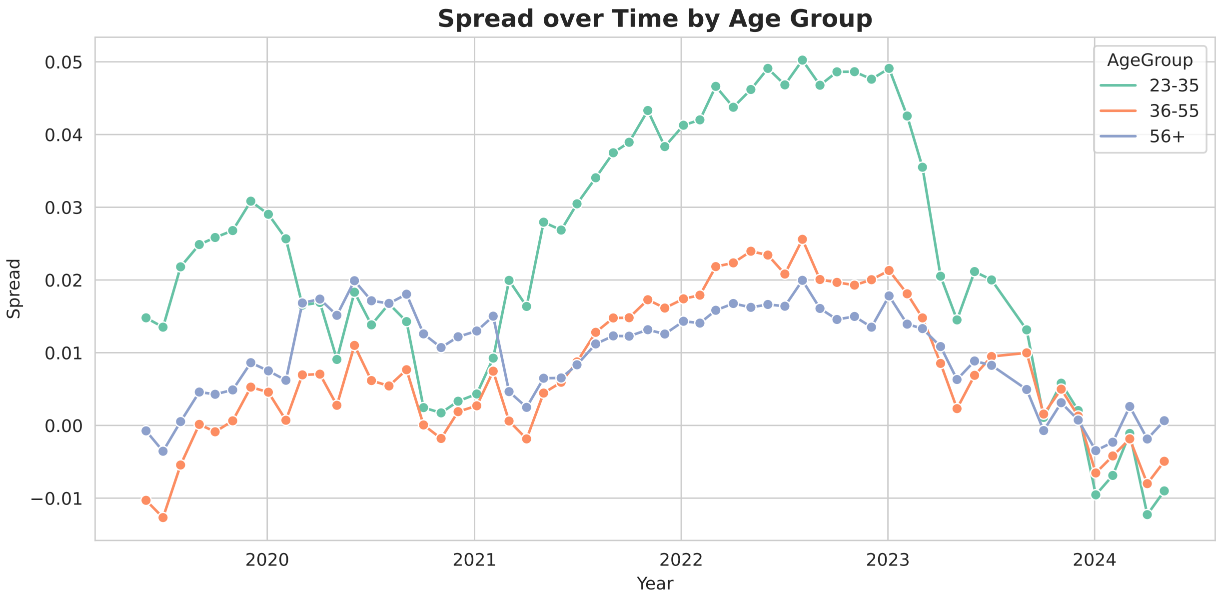

Post-Hoc Macroeconomic Adjustment of IFRS 9 Probability of Default Estimates

An Exploratory Study of Norion Bank's Swedish Unsecured-Loan Portfolio

Master's thesis in Engineering Mathematics and Computational Science

ADAM JOHANSSON

DEPARTMENT OF MATHEMATICS

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2025
www.chalmers.se

Master's thesis 2025

Post-Hoc Macroeconomic Adjustment of IFRS 9 Probability of Default Estimates

An Exploratory Study of Norion Bank's Swedish Unsecured-Loan Portfolio

ADAM JOHANSSON

Department of Mathematical Sciences
Chalmers University of Technology
Gothenburg, Sweden 2025

Post-Hoc Macroeconomic Adjustment of IFRS 9 Probability of Default Estimates
An Exploratory Study of Norion Bank's Swedish Unsecured-Loan Portfolio
Adam Johansson

Supervisor: Holger Rootzén, Department of Mathematical Sciences at Chalmers
Supervisor: Timmy Przybyla, Team Lead Data Scientist at Norion Bank
Examiner: Serik Sagitov, Department of Mathematical Sciences at Chalmers

Master's Thesis 2025
Department of Mathematical Sciences
Engineering Mathematics and Computational Science
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Observed spread $s_t^{(\mathrm{Age},\cdot)}$ over time, segmented by age group. Each line shows the monthly difference between observed and predicted default rates for a given borrower segment. A positive spread indicates that the current model underestimates default risk for that group.

Typeset in LaTeX
Gothenburg, Sweden 2025

Post-Hoc Macroeconomic Adjustment of IFRS 9 Probability of Default Estimates
An Exploratory Study of Norion Bank's Swedish Unsecured-Loan Portfolio
Adam Johansson
Department of Mathematical Sciences
Chalmers University of Technology

Abstract

This thesis examines whether post-hoc macroeconomic adjustments can explain - and potentially reduce - the spread between Observed 12-month Default rates ($\mathrm{OD}^{12\mathrm{m}}$) and the Probability of Default estimates-model ($\mathrm{PD}^{12\mathrm{m}}$) used by Norion bank for unsecured consumer loans in Sweden. The analysis cover June 2019 to February 2024, and consists of individual contracts that are aggregated to monthly-averages, a total of 56 months. We analyse the total portfolio in its entirety, as well as segmented on borrower characteristics like age, mortgage status, and co-borrower.

A Linear Regression model using ordinary least square (OLS) is employed to estimate the spread, defined as the difference between realized $\mathrm{OD}^{12\mathrm{m}}$ and predicted $\mathrm{PD}^{12\mathrm{m}}$. Correlation filtering, ElasticNet regularization, VIF screening, are used as feature selection techniques. Where autocorrelation is detected in the residuals, a Cochrane-Orcutt AR(1) correction is optionally introduced. The model is assessed using standard metrics like MAE, R^2, adjusted R^2, and residual diagnostics.

For the total portfolio, three macro variables—a 24-month change in the trade-weighted krona (KIX), a 6-month change in household confidence, and a 12-month change in unemployment - explains the spread. While out-of-sample test errors remain low, MAE comparison reveals overfitting. Segment analysis shows that models for young borrowers, older borrowers, and non-mortgage holders capture the patterns of observed spread, whereas middle-aged borrowers, mortgage holders, and single borrowers exhibit severe overfitting and unstable macroeconomic relationships.

The results indicate that linear macroeconomic overlays can add interpretive value and support $\mathrm{PD}^{12\mathrm{m}}$-model monitoring, but they are not yet stable enough for direct IFRS 9 adjustments. Key limitations include the short time series, structural changes in the macroeconomic landscape around COVID-19 time, and potential macro leakage already embedded in borrower-level $\mathrm{PD}^{12\mathrm{m}}$-models . Future research can explore non-linear models, rolling training windows, and direct macroeconomic integration into $\mathrm{PD}^{12\mathrm{m}}$ estimation.

Keywords: credit risk, IFRS 9, probability of default, expected credit loss, macroeconomic adjustment, regression, time series, segmentation, residual diagnostics, model validation

Acknowledgements

Curiosity has guided my academic journey at Chalmers which now ends with the completion of this thesis.

My sincere thanks to the entire Data Science team at Norion Bank for opening your doors to me and for all the *Cart*-games:

- Timmy Przybyla – for patient supervision, guidance, and encouragement.
- Johan – for sharing your earlier work on the topic and many helpful pointers.
- Mattias – for occasional, much-needed, focus-shifting small talk.
- João, my desk-neighbour – for daily laughs and a sharp eye in the final read-through.

I am grateful to my academic supervisor, Holger Rootzén, for giving so much of your time, insight, and good humour to this project. Also, thank you to examiner Serik Sagitov.

To every teacher, class-mate, and friend who shaped my years at Chalmers – thank you. A special nod to the "ENM-gang", and to Samuel for steadfast support from afar this final year.

To my family, whose unconditional support and encouragement made my studies at Chalmers possible – words are not enough. Finally, my thoughts and gratitude go to those who are no longer here to see the end of what *they* once helped to set in motion.

Thank you.

Adam Johansson, Gothenburg, June 2025

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis, listed in alphabetical order:

AR(1)	Autoregressive model of order 1
BG	Breusch–Godfrey test
BP	Breusch–Pagan test
CV	Cross-validation
DW	Durbin–Watson test
EAD	Exposure at Default
ECL	Expected Credit Loss
IFRS 9	International Financial Reporting Standard 9
LGD	Loss Given Default
MAE	Mean Absolute Error
OD	Observed Default
OLS	Ordinary Least Squares
PD	Probability of Default
VIF	Variance Inflation Factor

Nomenclature

Below is the nomenclature of indices, sets, parameters, and variables that have been used throughout this thesis.

Indices

t	Time index (monthly resolution)
l	Index for individual loan contracts
j	Index for features or macroeconomic variables

Sets

$\mathcal{M}^{(\cdot,\cdot)}$	Segment-specific portfolio
$\mathcal{M}^{\text{Total}}$	Total portfolio (all contracts)

Parameters

t_0	Valuation date (start of prediction horizon)
$t_{\max}$	Maximum number of months in the lifetime ECL horizon (typically 120)
n	Number of observations
p	Number of predictors
ρ	Autoregressive coefficient (AR(1))
λ	Regularization strength (Elastic Net)
α	Elastic Net mixing parameter (balance between L1 and L2)

Variables

$PD_{t_0,k}$	Probability that the contract defaults in bucket $[t_0+k-12,\ t_0+k)$, given survival up to that point
$\mathrm{PD}_{l,t}^{12m,\mathrm{Norion}}$	12-month probability of default for contract l at time t, predicted by Norion Bank's model
$\mathrm{PD}_{t}^{12m,\mathrm{Norion}}$	Used as notation for the current model for 12-month probability of default at Norion
$\overline{PD}_t^{\text{12m, Norion}}$	Portfolio-level average 12-month probability of default at t by Norions model
Adjusted $\overline{\mathrm{PD}}_t^{12m}$	Adjusted 12-month PD prediction at portfolio level (after adding $\hat{s}_t$)
$\overline{OD}_t^{12m}$	Observed default rate within 12 months after t
$EAD_{t_0,k}$	Expected exposure at default during month k after t_0
$LGD_{t_0,k}$	Expected loss given default during month k after t_0
$df(t_0, t_0+k)$	Discount factor from future month k back to valuation date t_0
s_t	Spread at time t (difference between observed and predicted default rate)
$\hat{s}_t$	Predicted spread at time t from the macro model
e_t	Residual at time t (model error)
X_t	Macroeconomic variable at time t
$X_t^{lag_i}$	Lagged feature of macro variable X
$X_t^{\Delta n}$	Delta feature (change over n months)
$X_t^{roll_k}$	Rolling average of macro variable X over k months
R^2, R^2_{adj}	Coefficient of determination and adjusted R^2

Contents

List of Figures

List of Tables

1
Introduction

This chapter presents the background, problem statement, purpose and research questions of the thesis. Credit risk, IFRS 9, modeling of PD^{12m} and the ECL-model are defined. Additionally, limitations, an overview of the methodology, and structure are presented.

1.1 Background

Banks and financial institutions play a central role in modern economy, facilitating growth through providing businesses and individuals with essential credit. For the general health of an economy and financial sector in particular, managing credit portfolios by assessing their risk, and mitigating the risk that a borrower fails to repay their debt is key.

Credit risk is a fundamental aspect of financial risk management, representing the potential loss a lender may incur if a borrower fails to meet their financial obligations. It can arise from borrower-specific factors, like income level and credit payment history, as well as macroeconomic conditions, such as interest rates, unemployment levels, and inflation. To avoid losing money, managing this credit risk is a central determinant of both the stability and profitability of a bank.

When a bank extends credit, it records a liability on their balance sheet, while offsetting an equivalent asset in the form of expected cashflow from the borrower. If the loan is repaid in full, the asset and the liability cancel out. However, if the borrower defaults, the bank makes an economic loss. Hence, its of utmost importance to have rigorous credit risk assessment and mitigation strategies to deal with default risks.

There are two main types of loans; (i) Secured loans where the loan is backed by collateral, and (ii) Unsecured loans that lack collateral meaning that, in the event of default, banks cannot recover their loss by seizing assets of the counter-party. Unsecured loans are inherently riskier, making it clear that robust models for predicting the probability of default, and rejecting customers with too high of a risk, are needed to avoid losses. Without adequate credit risk management and models, banks may face big losses, which could impact their financial stability. rejecting risky customers and identifying when existing customers are struggling are important and prioritized tasks of a healthy bank.

Norion Bank, is a niche financial institution in Sweden, and provides unsecured consumer loans ("blancolån") to the Swedish market under its brand *Collector*. Their portfolio of unsecured consumer loans will be the focus of this study. Credit risk management is particularly important for Norion due to the lack of collateral on these loans. Unsecured loans rely solely on the borrowers creditworthiness and repayment capacity, and when an eventual default occurs, rely on a debt collection agency to get their money back. With unsecured loans, banks are left with limited options for recovering the outstanding debt, making accurate default risk assessment and proper risk mitigation strategies essential.

To manage this risk, banks (and Norion) follow a structured approach based on the Expected Credit Loss (ECL) framework, which is required under the IFRS 9 accounting standard [2][1]. IFRS 9 requires banks to estimate their expected credit losses in advance and set aside capital reserves accordingly, called provisioning. This ensures that banks are financially prepared for potential defaults and that their financial statements accurately reflect the underlying risks in their loan portfolios.

Each month, t, for every active (i.e, not yet defaulted) loan, a model generates a predicted probability that the borrower will default on its loan within the next 12 months. In this thesis, Norion Banks probability of default model is denoted $\mathrm{PD}^{\text{12m, Norion}}$, the estimate in general is referred to as PD^{12m}.

One year later, at time $t+12$, the actual outcome is observed - either the borrower defaulted, or it did not. By comparing the average predicted $\mathrm{PD}_t^{\text{12m,Norion}}$, (denoted $\overline{\mathrm{PD}}_t^{\text{12m,Norion}}$) of all contracts in a portfolio on a given month, with the actual ratio of defaults observed one year later ($\overline{\mathrm{OD}}_t^{\text{12m}}$), we can evaluate the predictive accuracy of the current $\mathrm{PD}_t^{\text{12m, Norion}}$-model at Norion Bank. Note that the default can happen anytime during the 12-month prediction-horizon, but its first at time $t+12$ that we can calculate $\overline{\mathrm{OD}}_t^{\text{12m}}$.

For this study, we define, for each month t, the Spread, s_t, as the difference between these quantities:

$$s_t = \overline{\mathrm{OD}}_t^{\text{12m}} - \overline{\mathrm{PD}}_t^{\text{12m, Norion}} \tag{1.1}$$

Ideally, the spread should be close to zero. In that case, the predicted default risk aligns with the actual outcomes on an aggregated portfolio-level. A positive spread ($s_t > 0$) means that the current $\mathrm{PD}^{\text{12m, Norion}}$-model underestimated the true default rate, suggesting that the model was too optimistic. A negative spread ($s_t < 0$) suggest overestimation. Underestimation of default risk may result in insufficient capital reserves, exposing the bank to unexpected credit losses. Overestimation leads to excessive provisioning, tying up capital unnecessarily. Both are undesirable, but overestimation is generally considered better since its a regulatory-safe option. The spread concept is further developed in Section 2.2.

[1] `https://www.ifrs.org/issued-standards/list-of-standards/ifrs-9-financial-instruments/`

1.2 Problem Statement

One of the key components when calculating expected credit loss (ECL) is the Probability of Default (PD) explained above. A precise estimation of PD enables banks to allocate adequate credit loss provisions — impacting both the income statement and capital requirements under regulatory frameworks such as IFRS 9. In this study we focus on the 12-month probability of default, PD^{12m}.

Accurate predictions of PD^{12m} are challenging in some aspects. The current $\text{PD}_t^{\text{12m, Norion}}$-model at Norion Bank is proprietary, but in general, such models are based on borrower and contract-level characteristics. At the moment, Norion Bank's model does not explicitly take into consideration current, past or predicted future macroeconomic conditions. Yet, it is intuitive and well-established in the literature [3], that macroeconomics influence the default risk. Default risks can be sensitive to shifting macroeconomic conditions, such as unemployment, inflation, and interest rates. Thus, there might be a way to improve the current $\text{PD}_t^{\text{12m, Norion}}$-model using macroeconomic indicators.

The argument is as follows: since the current $\text{PD}_t^{\text{12m, Norion}}$-model does not explicitly include macroeconomic indicators, some of the discrepancy between the predicted and observed default may be explained using macro indicators. If we can model this discrepancy, spread, using macroeconomic indicators, then a more accurate *adjusted* $\overline{\text{PD}}^{\text{12m}}$-model can be achieved, and ultimately enhance the performance of credit risk models at Norion Bank. Specifically, the adjusted prediction would look like:

$$\text{Adjusted } \overline{\text{PD}}_t^{12m} = \overline{\text{PD}}_t^{\text{12m, Norion}} + \hat{\text{s}}_t \tag{1.2}$$

where $\hat{\text{s}}_t$ is this report's model's estimation of the spread, and $\overline{\text{PD}}_t^{\text{12m, Norion}}$ is Norion Bank's monthly average prediction of all contracts.

1.3 Purpose and Research Questions

The purpose of this thesis is to improve Norion Bank's $\text{PD}^{\text{12m, Norion}}$ -model predictions by modeling the spread, s_t using macroeconomic factors. With the more specific research questions as a starting point:

- To which extent can macroeconomic variables account for the gap between observed default rates and predicted probabilities?
- How does borrower segmentation (age, co-borrower status, mortgage status) influence sensitivity to macroeconomic changes?
- To what degree can macroeconomic indicator–based post-hoc adjustments enhance predictive accuracy compared to the current $\overline{\text{PD}}_t^{\text{12m, Norion}}$ prediction?

1.4 Scope and Limitations

This thesis focus on unsecured consumer loans in the Swedish market, specifically covering Norion Bank's entire portfolio of private consumer loans. The analysis excludes all forms of secured loans like mortgages, as they differ in both structure, risk, and are not offered by Norion.

The analysis is both conducted at the total portfolio level and across several borrower segments, including age groups (23-35, 36-55, 56+), mortgage-holder status (Yes, No, Unknown), and co-borrower (Yes/No). These segments were made with the idea that different borrowers might be more or less sensitive to specific macroeconomic factors, and to investigate how the current model performs across different segmentation.

The study aggregates individual loan data into 56 monthly portfolio-level averages. The start date was chosen to reflect a stable period on the current $\text{PD}^{\text{12m, Norion}}$ -model, and to avoid inconsistencies introduced by changes in previous model versions. However, this relatively short time window introduces some limitations. Especially; (i) the limited number of data points, which restricts model options, and (ii) we are exposed to a period of structural breaks in economic conditions as the effect of the COVID-19 pandemic influences the model's fit. Macroeconomic data is obtained from public sources and aligned to monthly frequencies. They were chosen to capture as broad a picture ass possible of the macroeconomic conditions.

This study initially set out with the ambition of developing a model potentially ready for implementation. However, as the work progressed, it became clear that the problem was more complex than expected. Hence, the focus was shifted towards a more exploratory analysis. The thesis instead serves as a feasibility study, investigating whether post-hoc macroeconomic adjustments can help explain the prediction-error in the current $\text{PD}_t^{\text{12m, Norion}}$-model.

1.5 Methodology (Brief Overview)

The methodological approach in the thesis follows a structured modeling framework designed to identify the most influential macroeconomic indicators on the spread. We build a separate models for the total portfolio, and for each of the segmented portfolios. It starts off using correlation measures as a tool for initial filtering, followed by Elastic Net regularization for feature selection, and VIF filtering to address multicollinearity issues.

Linear Regression models are estimated using OLS, with optional AR(1) correction applied via Cochrane–Orcutt method when residuals show autocorrelation. The performance of the model is assessed using standard metrics like R^2, adjusted R^2, and MAE, alongside diagnostic tests for the residuals.

1.6 Use of AI Tools

OpenAI's ChatGPT, were used in this thesis to support writing, clarify some technical concepts and to suggest the structure of the report. Specifically, ChatGPT was used to:

- Refine the clarity, grammar, and flow of text that was already drafted. This was especially used in the methodology, results, and discussion chapters. All text have gone through the AI to look for typos and inconsistency with symbols and notations.
- Analyze the results from coding and give suggestions on interpretation.
- Provide formatting for tables, equations, and figure captions to improve visual consistency.

No text or code was accepted without review. All AI-generated text content was edited, and adjusted to reflect my own understanding. Final responsibility for the content, analysis, and conclusions rests entirely with the author.

1.7 Thesis Structure

The thesis is divided into the following chapters:

Chapter 2 (Theory): This chapter reviews relevant theories, and concepts. We start off by introducing the IFRS 9 framework and ECL-model. The macroeconomic indicators are presented, and how we feature engineer them to account for lagged effects. Lastly, the mathematical theory is outlined.

Chapter 3 (Methodology): In the methodology chapter we write about how the data is gathered and handled. We present the data using tables and plots. Additionally we introduce the modeling framework for the modeling part. Limitations and data reliability are also addressed.

Chapter 4 (Results): Here we present empirical findings from the modeling. Plots, tables and diagnostics are used to describe the results. The chapter is divided so that each portfolio and borrower segmentation is presented in a clear manner.

Chapter 5 (Analysis and Interpretation): This chapter interprets the findings from the Results chapter. Firstly, we analyse the observed spread and how it differs for the segmented portfolios. Secondly, we look at the modeling performance of the models across the segmentation. Lastly, we interpret the meaning of the selected macroeconomic features for each model.

Chapter 6 (Discussion): Here, we reflect on the findings, implications, and limitations of the study. Some takeaways from the segmentation, practical implication for Norion Bank and its $\text{PD}^{\text{12m, Norion}}$-modeling. Lastly, we suggest directions for future research.

2
Theory

This chapter presents the theoretical foundation for the thesis. Firstly, we introduce the IFRS 9 framework and especially the ECL-model. Secondly, we present the macroeconomic indicators relevant to the study. Thirdly, we present the mathematical concepts needed for the result and analysis therof.

2.1 The IFRS 9 Framework

The International Financial Reporting Standard 9 (IFRS 9) is an accounting standard issued by the International Accounting Standards Board (IASB) that governs how financial institutions recognize, measure, and thesis financial instruments. IFRS 9 was introduced to replace the previous IAS 39 standard, which was criticized for its delayed recognition of credit losses, especially during financial crisis [4]. The primary goal of IFRS 9 is to create a more forward-looking, risk-sensitive framework for recognizing and provisioning for Expected Credit Losses (ECL).

Under IAS 39, banks and financial institutions followed an incurred loss model, meaning they only recognized loan losses after a borrower defaulted or when there was clear evidence of impairment. This approach resulted in "too little, too late" provisioning, as losses were often recorded only after the financial damage had already occurred. This often lead to under-provisioning during economic booms. During economic downturns, under-provisioning led to sudden spikes in credit losses, causing volatility in financial statements and increasing financial instability.

To address these shortcomings, IFRS 9 introduced the Expected Credit Loss (ECL) model, which requires financial institutions to estimate and provision for potential credit losses in advance, even if no default has yet occurred. This ensures that banks and lenders maintain adequate reserves for potential losses, and makes the financial system more resilient and transparent.

2.1.1 Expected Credit Loss (ECL-) model

The Expected Credit Loss (ECL) model under IFRS 9 estimates the present value of credit losses over a defined time frame. The ECL for a contract l, at time t_0, and time frame TF, is defined as:

$$ECL_{t_0}^{\mathrm{TF}} = \sum_{k \in T} PD_{t_0,k}\, EAD_{t_0,k}\, LGD_{t_0,k}\, df(t_0, t_0 + k),$$

where

- $T = \{12\text{m}, 24\text{m}, \dots, t_{\max}\}$ is the set of 12-month time buckets extending to contractual maturity;
- $PD_{t_0,k}$ is the probability that the loan defaults during the bucket $[t_0 + k - 12,\ t_0 + k)$, conditional on no earlier default;
- $EAD_{t_0,k}$ is the expected exposure at default in that bucket;
- $LGD_{t_0,k}$ is the loss-given-default fraction for that bucket;
- $df(t_0, t_0 + k)$ is the discount factor that brings the expected loss in bucket k back to present value at t_0.

IFRS 9 divides the provisioning methodology into three stages with different requirements on the scope of provisioning based on the performance of the loan.

1. Stage 1: These are loans where there have been no significant increase in credit risk since the credit was issued. The bank expects the borrowers to continue to meet their payment obligations. In this stage, the ECL is calculated over a 12-month horizon, $T = \{12m\}$ (months after t_0).

2. Stage 2: If a borrower experience a significant deterioration in creditworthiness, their loan is moved to Stage 2. What constitutes significant deterioration is not defined by IFRS 9, but is left to the bank. These loans are not yet in default, but their risk has increased enough so that the bank needs to estimate credit losses over the entire remaining lifetime of the loan. The time frame spans the remaining lifetime of the loan, and $T = \{12m, 24m, ..., t_{\max}\}$.
3. Stage 3: Here, the credit is considered impaired, which means that $PD = 1$. Usually, the bank has determined that the borrower is unlikely to repay their loan without some form of intervention (e.g., debt collection), or the borrower is more than 90 days past due on a payment. There is no uncertainty around EAD, and LGD will now show the best estimate of credit losses.

An overview of the stage modeling can be seen in table 2.1.

Stage	**Credit Risk Status**	**ECL Horizon**
1	No significant increase in credit risk	12 months
2	Significant deterioration in credit risk	Lifetime
3	Credit impaired (PD = 100%)	Lifetime (defaulted)

Table 2.1: Summary of IFRS 9 Staging and Expected Credit Loss Calculations

A critical requirement under IFRS 9 is that PD (and LGD) estimates must incorporate macroeconomic factors [5]. Since credit risk is not static and reasonably fluctuates with economic conditions, relying solely on debtor-specific historic data might fail to capture potential shifts in default risks. Rising unemployment and interest rates may reduce borrowers' ability to repay a debt; conversely, declining unemployment and overall economic growth, could reduce default risk. How macroeconomic variables should be incorporated is not decided in IFRS 9, but is up to the bank itself.

2.2 Probability Of Default (PD), Observed Default (OD), and Spread

Probability of Default (PD) is a fundamental measure in ECL-modeling, representing the likelihood that a borrower or loan contract will default within a defined time horizon - typically within the next 12 months (PD_t^{12m}), or for the remainder of the credit term ($\mathrm{PD}_t^{\mathrm{Lifetime}}$). As stated earlier we focus primarily on the 12-month probability of default PD^{12m}.

Each month t, a 12-month probability of default $\mathrm{PD}_{l,t}^{12m}$, is generated for every active loan l. A PD_t^{12m}-model is usually based on borrower characteristics, historical credit performance, and other relevant data at the time of prediction (the valuation date).

Formally, at time t, for an individual loan contract l, the predicted probability of default within the next 12 months is defined as:

$$PD_{l,t}^{\mathrm{12m}} = P(\tau_l \leq t + 12 \mid \mathcal{F}_t, \tau_l > t) \tag{2.1}$$

where, τ_l is the default time for loan contract l, $\mathcal{F}_t$ represents all available information at valuation date t, including all knowledge about the customer, their payment history etc.

Every month, all active contracts l receive a new $PD_{l,t}^{\text{12m, Norion}}$ estimate from the current $\mathrm{PD}_t^{\text{12m, Norion}}$ model. In turn, these predictions are aggregated at the portfolio level, $\overline{\mathrm{PD}}^{\text{12m, Norion}}$, to evaluate the performance of the bank's $\mathrm{PD}^{\text{12m, Norion}}$ model. Specifically, the aggregated portfolio-level prediction $\overline{\mathrm{PD}}_t^{\text{12m, Norion}}$ at month t is calculated as the mean across all active contracts:

$$\overline{\mathrm{PD}}_t^{\text{12m, Norion}} = \frac{1}{N_t} \sum_{l=1}^{N_t} \mathrm{PD}_{l,t}^{\text{12m, Norion}} \tag{2.2}$$

where N_t denotes the number of active loan contracts in the portfolio at time t.

To evaluate the accuracy of a bank's current $\mathrm{PD}^{\text{12m, Norion}}$-model, we can compare the aggregated portfolio-level prediction $\overline{\mathrm{PD}}_t^{\text{12m, Norion}}$ to the portfolio-level observed defaults. For each valuation date t, we count whether a loan contract defaulted within the following 12-months (i.e., between month t and $t + 12$).

Formally, the aggregated observed default rate, $\overline{\mathrm{OD}}_t^{12m}$, at month t is calculated as:

$$\overline{\mathrm{OD}}_t^{12m} = \frac{|\{l \in R_t : \tau_l \in [t, t+12]\}|}{|R_t|} \tag{2.3}$$

where, τ_l is the observed default time for loan l, and R_t is the risk set at time t, i.e., the set of loans that are active and not yet defaulted as of month t.

Thus, $\overline{\mathrm{OD}}_t^{12m}$, serves as the observed benchmark against which the current model is measured and evaluated. It effectively describes the risk at the portfolio-level.

2.2.1 Spread

For the purpose of improving portfolio-level predictions, we define a key modeling target, called "Spread," which quantifies the difference between the actual observed default rate $\overline{\text{OD}}_t^{12m}$ and the portfolio-level prediction from Norion's current model, $\text{PD}^{\text{12m, Norion}}$.

Specifically, the Spread at month t is defined as:

$$\text{s}_t = \overline{\text{OD}}_t^{12m} - \overline{\text{PD}}_t^{\text{12m, Norion}} \tag{2.4}$$

The Spread, s_t captures the error of the current $\text{PD}^{\text{12m, Norion}}$ model, highlighting potential systematic biases or inaccuracies in the bank's predictive models. A positive Spread ($\text{s}_t > 0$) means that the observed portfolio-level default rate was higher than predicted, i.e., the current PD^{12m}-model underestimated risk at that valuation date. A negative Spread ($\text{s}_t < 0$) shows an overestimation of risk.

The current PD^{12m}-model at Norion Bank does not explicitly incorporate macroeconomic factors, instead a macro adjustment is applied on ECL level.

It's important to highlight the inherent 12-month delay between the prediction at time t and the realization of the default outcomes at time $t+12$. Specifically, while the $\text{PD}_{l,t}^{\text{12m, Norion}}$, for contract l is predicted at month t, the observed default outcome for each contract is known first at time $t+12$. This results in a 12-month lag in the realization of the spread s_t. Therefore, the spread s_t, can only be determined retrospectively after this 12-month lag. As a result, model errors or bias may go undetected for up to a year. Making it even more crucial to have accurate predictive models.

By modeling the spread, s_t, as a function of macroeconomic indicators, we can adjust the prediction of the $\overline{\text{PD}}_t^{\text{12m, Norion}}$. Formally, if we succeed in accurately modeling the relationship between macroeconomic variables and the Spread, we can construct a predictive adjustment term, $\hat{\text{s}}_t$, which can be added to the original $\text{PD}^{\text{12m, Norion}}$-model:

$$\text{Adjusted } \overline{\text{PD}}_t^{12m} = \overline{\text{PD}}_t^{\text{12m, Norion}} + \hat{\text{s}}_t \tag{2.5}$$

This adjustment would close the gap between the predictions and the observed outcomes, improving the reliability and robustness of the bank's PD^{12m}-model, thereby enabling more accurate provisioning.

2.3 Macroeconomy

Macroeconomy refers to the overall structure, performance, and behavior of an economy at a national or global level. It covers broad economic indicators such as gross domestic product (GDP), inflation (KPI), unemployment rates, and interest rates, which collectively gives an indication of the well-being of an economy.

Since macroeconomic conditions impact income levels, financial stability, and borrowing capacity of the public, they play a role in determining credit risk and default probabilities in banking and lending sectors. Understanding how macroeconomic indicators impact default risk is therefore essential for financial institutions in general, and for Norion Bank in particular. This section will introduce the macroeconomic indicators used in the study.

2.3.1 Macroeconomic indicators of interest

Inflation and KPIF

Inflation is the rate at which the general price level of goods and services increases over time. The result of inflation is a reduction in the purchasing power of money. It is influenced by multiple factors, like demand and supply imbalances, and monetary policies set by central banks. In Sweden, inflation is measured using the Consumer Price Index (CPI, 'KonsumentPrisIndex' or KPI in Swedish), which is calculated by tracking price changes of a fixed basket of goods and services over time [7]. The basket contains, among other things, cost of food, housing, and transportation. The goods are weighted based on their share of total consumer spending. The year-over-year KPI figure reported for a given month, such as March 2024, reflects cumulative price changes from March 2023 to March 2024. This means the inflation signal is inherently lagged. It captures price developments over the previous 12 months.

KPI is reported monthly and the index can be expressed relative to a base year. KPIF ('KonsumentPrisIndex med fast ränta'/'Consumer Price Index with Fixed Interest Rates) is an inflation measure that adjust the general KPI to remove the direct effects of interest rate changes. The KPIF indicator keeps interest rates constant to provide a more clear view of the underlying inflation trend. KPIF is the Riksbanks target variable to lead them in decisions to achieve their inflation target of 2%. KPIF is measured in current prices. An alternative is the KPIF Fixed Baseline, which compares changes to a base year.

Policy rate

The Policy rate ('styrräntan', previously 'Reporäntan') is the key interest rate set by the central bank of Sweden, the Riksbank [8]. It is the rate at which commercial banks can borrow money from the Riksbank.

When inflation is high, the Riksbank raises the policy rate, increasing the cost of borrowing money, resulting in lower consumption, and thereby reducing inflation. Whereas a lower policy rate stimulates the economy by making credit cheaper. The effect of the policy rate on the economy is often lagged, as it takes time for higher interest to impact consumptions and borrowing costs. From a credit risk perspective,the policy rate might influence default probabilities with a delay, as borrowers gradually experience higher debt. The policy rate is decided by the Executive Board of the Riksbank at predetermined meetings five times a year. If economic conditions change significantly, additional rate decisions can be made outside the scheduled meetings. The policy rate can be retrieved through the Riksbanks API.

Unemployment

Unemployment refers to the share of the labor force actively seeking but unable to find employment. High unemployment signals economic downturns, increasing financial stress and reduced household income. During high unemployment, borrowers usually experience a decline in their financial situation, increasing the risk for default. Therefore, high unemployment could be linked to higher credit risk [9], as borrowers' income decline, and thereby their ability to repay debt diminishes. Taking the unemployment level into consideration when dealing with default risk might therefore be of interest for banks.

Gross Domestic Product (GDP)

Gross Domestic Product (GDP) /('BruttoNationalProdukt' (BNP)) is the total monetary value of all finished goods and services produced within a country during a specific time period, usually in one year or a quarter. It can be expressed in nominal (current prices) or real (inflation-adjusted) terms. Real GDP accounts for price-level changes to reflect actual economic growth.

An increase in GDP indicates economic growth, and decreasing GDP signalizes economic contraction or recession. GDP helps policymakers and businesses evaluate the overall economic performance and stability. For banks, the trends of GDP are important because economic expansions typically lead to improved economy for borrowers, typically lower credit risk [10].

KIX-Index

The KIX Index ('KronIndeX') measures the Swedish krona's value against a weighted basket of currencies [11] from Sweden's major trading partners, reflecting the international competitiveness of Sweden. A higher KIX-index indicates a weaker krona, boosting the competitiveness of exports but making imports more expensive. Fluctuations in KIX can influence inflation, and consumer purchasing power, indirectly impacting credit risk. The KIX index can also be viewed as an indicator of how the global market perceives Sweden's economic stability and attractiveness. A weaker krona may also stretch household finances through more expensive travel and online consumption from abroad.

Economic sentiment from NIER

The National Institute of Economic Research (NIER) / ('Konjukturinstitutet') publishes several macroeconomic indicators that provide insight into Sweden's economic conditions [12]. These include:

- Economic Barometer, a composite indicator summarizing the overall economic confidence among households and businesses. Indicating trends in future economics and current conditions.
- Consumption Indicator, reflects household consumer patterns, trends, and are indicative of the willingness and ability to spend money.
- Business Barometer, aggregates sentiment from businesses across multiple sectors and industries, with the purpose of capturing business confidence regarding economic outlook and conditions.
- Retail Trade Indicator, measures the confidence and economic outlook specifically within the retail sector.
- Service Sector Indicator, reflects sentiment and performance in service-oriented businesses.
- Consumer Confidence Indicator, assesses the households perceptions of their personal economic situations and general economic outlook.
- Household Macro Index summarizes households overall perception on macroeconomic conditions, giving an overall insight into consumer confidence and potential future consumption behavior.

All these indicators reflect the state of the household economy, and by extension the financial stability of borrowers and their ability to manage debt obligations.

2.4 Mathematical Concepts, Modeling and Diagnostics.

This section outlines the mathematical methods and techniques used to analyze and model the relationship between macroeconomic indicators and Spread. It includes approaches like feature selection, correlation analyses, regularization techniques, and regression modeling, and model diagnostics.

2.4.1 Feature Engineering

To capture the delayed effect of macroeconomic variables on credit risk, one can derive a set of engineered features from the raw data [13]. The feature engineering process involves generating three distinct categories of transformations. For each macroeconomic indicator X_t, the following transformations are applied:

- Lagged Features, which capture delayed or lagged economic effects.
- Rolling Average Features, used to smooth short-term volatility and highlight medium-term trends.
- Delta Features, which quantify recent absolute changes, thus capturing momentum and trend shifts.

Lagged features capture the delay between macroeconomic events and their impact on credit risk. These variables can help the model capture delayed economic reactions that are often observed in credit market behavior. For each macroeconomic variable X_t, lagged features are defined as:

$$X_t^{lag_i} = X_{t-i}, \quad \text{where } i = 1, 2, 3, \ldots, k \tag{2.6}$$

Here, $X_t^{lag_i}$ is the lagged value of the macroeconomic variable X at time t with a lag of i months.We generate lagged variables up to a maximum lag length of $k = 30$, based on prior analysis at Norion Bank.

Rolling average (moving average) features reduce short-term noise and are aimed at helping to isolate meaningful economic signals over specified periods. These features might provide indications of stable or underlying trends by averaging out monthly volatility. In this thesis they are calculated as arithmetic means over fixed rolling windows of 3, 6, and 12 months:

$$X_t^{roll_k} = \frac{1}{k}\sum_{j=0}^{k-1} X_{t-j}, \quad k \in \{3, 6, 12\} \tag{2.7}$$

In this equation, $X_t^{roll_k}$ denotes the rolling average value of feature X at month t, calculated over the previous k months.

Delta features quantify absolute changes in macroeconomic indicators over different time horizons, capturing short-term and long-term shifts in economic conditions. They are defined as:

$$X_t^{\Delta n} = X_t - X_{t-n}, \quad n \in \{1, 3, 6, 12, 24\} \tag{2.8}$$

where $X_t^{\Delta n}$ represents the difference in the indicator's value from n months ago.

Additionally, a momentum shift feature is introduced:

$$X_t^{\Delta 12,24} = X_{t-12} - X_{t-24} \tag{2.9}$$

To be clear $X_t^{\Delta n}$ represents the absolute difference over the previous n months, and $X_t^{\Delta 12,24}$ captures the momentum shift from one year ago compared to two years ago.

2.5 Feature Selection Methods

Feature selection is a systematic procedure aimed at reducing the dimensionality of the dataset by identifying and retaining the most relevant predictors. This process is used to improve the model interpretability, reduce overfitting, and enhance prediction accuracy. The feature selection methodology in this study includes statistical measures of correlation, regularization techniques (ElasticNet), and multicollinearity checks (Variance Inflation Factor).

Correlation analysis measures the strength and direction of a *linear* relationship between two numerical variables [15]. Pearson's correlation coefficient, denoted by ρ, measures the linear association between two numerical variables, X and Y. It is defined as:

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}, \tag{2.10}$$

where $\text{Cov}(X, Y)$ is the covariance between X and Y, σ_X and σ_Y represent their respective standard deviations, and $\bar{X}$, $\bar{Y}$ are the sample means. Pearson's correlation coefficient ranges between -1 and $+1$, indicating a perfect negative and positive linear relationship, respectively, while 0 indicates no linear correlation. A negative linear relationship means that as one variable increases, the other decreases.

ElasticNet is a hybrid regularization technique that combines properties from both Ridge (L2-regularization) and LASSO (L1-regularization) [14]. ElasticNet balances feature selection (L1) and coefficient shrinkage (L2), addressing multicollinearity and simultaneously performing variable selection. The ElasticNet optimization problem is expressed as:

$$\min_{\boldsymbol{\beta}} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \left[\alpha \|\boldsymbol{\beta}\|_1 + \frac{(1-\alpha)}{2} \|\boldsymbol{\beta}\|_2^2 \right] \right\}, \tag{2.11}$$

where:

- $\mathbf{y}$ is the vector of observed outcomes.
- $\mathbf{X}$ is the matrix of predictor variables.
- $\boldsymbol{\beta}$ represents the vector of regression coefficients.
- $\|\boldsymbol{\beta}\|_1 = \sum_{j=1}^{p} |\beta_j|$ is the L1 norm.
- $\|\boldsymbol{\beta}\|_2^2 = \sum_{j=1}^{p} \beta_j^2$ is the squared L2 norm.
- $\lambda \geq 0$ controls the strength of the regularization penalty.
- $\alpha \in [0, 1]$ determines the balance between L1 and L2 regularization; $\alpha = 1$ corresponds to pure LASSO, and $\alpha = 0$ corresponds to pure Ridge regression.

In this thesis, we remove features with zero coefficients after Elastic Net.

Optimal ElasticNet coefficients are typically determined using cross-validation to select hyperparameters λ and α. Cross-validation (CV) is a statistical procedure used to evaluate the generalization capability of predictive models by partitioning the original dataset into subsets. It assesses how the results of a statistical model generalize to an independent data set. In K-fold cross-validation the dataset is randomly partitioned into K equally sized subsets (folds). The model is trained K times, each time using $K-1$ folds as the training set and the remaining fold as the test set.

Formally, let the data set $D = \{(x_i, y_i)\}_{i=1}^{N}$ be partitioned into K equally sized disjoint subsets $D_1, D_2, \ldots, D_K$. Then, for each fold $k \in \{1, 2, \ldots, K\}$, the cross-validation error is calculated as:

$$\text{CV Error} = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{|D_k|} \sum_{(x,y) \in D_k} L\left(y, \hat{f}^{-k}(x)\right), \tag{2.12}$$

where, D_k is the k-th subset used for validation, $\hat{f}^{-k}(x)$ represents the model trained on all subsets except D_k, $L(y, \hat{y})$ is a loss function measuring the discrepancy between actual y and predicted $\hat{y}$. The optimal model parameters are selected based on the result of minimization of the cross-validation error.

The Variance Inflation Factor (VIF) quantifies the severity of multicollinearity among explanatory variables in a regression model. Multicollinearity occurs when predictors are highly linearly correlated, causing instability in coefficient estimates when using Ordinary Least Square (OLS). VIF measures how much the variance of an estimated regression coefficient increases due to collinearity among explanatory variables. VIF of X_j is defined as:

$$\text{VIF}_j = \frac{1}{1 - R_j^2}, \tag{2.13}$$

where R_j^2 is the coefficient of determination obtained from regressing the explanatory variable X_j on all other explanatory variables.:

$$X_j = \beta_0 + \sum_{\substack{i=1 \\ i \neq j}}^{p} \beta_i X_i + \epsilon_j. \tag{2.14}$$

where, R_j^2 measures the proportion of variance in X_j explained by other predictors, a high R_j^2 (close to 1) implies severe multicollinearity, resulting in a high VIF.

To mitigate multicollinearity, features with a Variance Inflation Factor (VIF) exceeding 10 were excluded. This threshold is commonly used as a rule of thumb [16], indicating that a variable is highly collinear with others (i.e., $R_j^2 > 0.9$). Although, the threshold is quite arbitrary and up to discussion.

2.6 Linear Regression Modeling

Regression modeling is a technique used to quantify the relationship between independent variables (predictors/covariates/features/explanatory variables) and a dependent variable (target). One linear regression model is Ordinary Least Squares (OLS). It is a method for estimating the parameters of a linear regression model. The OLS method seeks to minimize the sum of squared residuals (errors) between the observed values and the predicted values. Given a dataset with n observations and p predictors, the linear regression model is defined as:

$$Y = X\beta + \epsilon, \tag{2.15}$$

here, Y is a $n \times 1$ vector containing the observed dependent variable values (target), X is the $n \times p$ matrix of independent variables, β is the $p \times 1$ vector of unknown regression coefficients, ϵ is the $n \times 1$ vector of residuals (errors).
The OLS estimator for β is given by:

$$\hat{\beta} = (X^T X)^{-1} X^T Y. \tag{2.16}$$

There are five key assumptions of OLS [17]:

1. Linearity: The relationship between independent variables and the dependent variable is linear, or the relationship is expected to be linear.
2. Independence: The residuals (ϵ_i) are independently and identically distributed (i.i.d).

3. Homoscedasticity: The residuals, (ϵ_t), have constant variance.
4. No perfect multicollinearity: No independent variable is a perfect linear combination of other independent variables.
5. Normality of Residuals: is assumed to follow a normal distribution with mean zero and constant variance.

If these assumptions are violated, OLS estimates can be inefficient, biased, or misleading.

2.6.1 Model Diagnostics

Evaluating the validity of a regression model requires diagnostic tests to ensure that the underlying assumptions hold. This section outlines the statistical techniques used to detect multicollinearity, autocorrelation (trends in residuals), heteroscedasticity, and residual normality.

The goodness-of-fit of an OLS model can be evaluated using the R-squared (R^2) statistic, which measures the proportion of the variance in the dependent variable that is explained by the independent variables:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}. \tag{2.17}$$

where y_i represents the actual observed value, $\hat{y}_i$ represents the predicted values, $\bar{y}$ is the mean of the observed values.

While R^2 provides a measure of how well the model explains the variance in the data, it has some limitations:

1. It always increases when more variables are added, even if they do not improve model performance [18].
2. It does not account for the number of predictors used.

An alternative, to address these limitations, is the Adjusted R-squared (R^2_{adj}). It adjusts for the number of predictors in the model, and is given by:

$$R^2_{\text{adj}} = 1 - \left(\frac{(1 - R^2)(n-1)}{n-p-1} \right), \tag{2.18}$$

where, n is the number of observations, p is the number of predictors. If a new predictor does not add explanatory power, R^2_{adj} will decrease.

2.6.2 Residual Diagnostics

Residuals, defined as the difference between observed values and model predictions, should exhibit properties such as mean zero, constant variance, and independence. Formally, for an estimated regression model:

$$e_t = y_t - \hat{y}_t, \tag{2.19}$$

where e_t is the residual of observation t, y_t is the observed value, and $\hat{y}_t$ is the predicted value. Normality and homoscedasticity tests (described below) are performed to ensure the validity of the regression model.

MAE quantifies model error by measuring the average magnitude of errors in the model's predictions [19]. It is defined as:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| y_i - \hat{y}_i \right| , \tag{2.20}$$

where y_i denotes the observed value, $\hat{y}_i$ the predicted value, and n the number of observations.

The Breusch–Pagan (BP) [23] test evaluates whether the residual variance in a regression model is constant (homoskedastic) or varies systematically with the independent variables (heteroskedasticity). Under the null hypothesis, the residuals (e_t) have constant variance:

$$\text{Var}(e_t \mid X_t) = \sigma^2.$$

To test this, the squared residuals from the original model are regressed on the original set of predictors:

$$e_t^2 = \alpha + \sum_{j=1}^{p} \gamma_j X_{j,t} + u_t, \tag{2.21}$$

where e_t^2 is the squared residual from the main regression model, $X_{j,t}$ are the predictors, and u_t is a new error term. The test statistic is:

$$LM = nR^2 \sim \chi_p^2,$$

where R^2 is the coefficient of determination from the auxiliary regression, p is the number of predictors. A significant result indicates heteroskedasticity.

The Shapiro–Wilk test [20] assesses whether a sample (e.g., residuals) comes from a normal distribution. It uses a weighted correlation between ordered residuals and expected normal order statistics:

$$W = \frac{\left(\sum_{t=1}^{n} a_t e_{(t)} \right)^2}{\sum_{t=1}^{n} (e_{(t)} - \bar{e})^2}, \tag{2.22}$$

where $e_{(t)}$ are the ordered residuals, $\bar{e}$ is the sample mean of residuals, and a_t are constants derived from the expected values of order statistics of a standard normal distribution. The hypotheses are:

- H_0: Residuals are normally distributed.
- H_1: Residuals deviate from normality.

A p-value less than 0.05 leads to rejecting H_0, suggesting non-normality.

2.7 Time Series Modeling

Residuals from regression models should ideally be white noise, meaning they exhibit no autocorrelation and have constant variance. However residuals often display temporal dependencies, meaning that past values influence current values. Let's be clear that we investigate autocorrelation on the residuals, not the observations. If autocorrelation is detected—particularly of the AR(1) type—it violates one of the OLS assumption. Therefore, this section not only introduce tests for autocorrelation, but also shows a correction method when first-order autocorrelation is present.

The autoregressive model of order 1 (AR(1)) of the residuals is the simplest time-series model used to assess whether the residuals follow a systematic pattern. It assumes that the current residual value e_t depends on its immediately preceding value e_{t-1} and a stochastic error term:

$$e_t = \phi e_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2) \tag{2.23}$$

where, e_t represents the residual at time t, ϕ is the autoregressive coefficient that measures the dependency on past residuals, ϵ_t is a white noise error term with mean zero and constant variance σ^2. If $\phi = 0$, there is no autocorrelation, indicating that residuals are independent. If $\phi > 0$, residuals exhibit positive autocorrelation (trending behavior). If $\phi < 0$, residuals show negative autocorrelation (mean-reverting behavior).

The Durbin-Watson (DW) test [22] detects the presence of first-order autocorrelation in the residuals, e_t, meaning there are trends in the residuals. The test statistic is computed as:

$$DW = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}, \tag{2.24}$$

where e_t are the residuals. The null hypothesis of the Durbin–Watson test is that the residuals exhibit no first-order autocorrelation, i.e., $\text{Cov}(e_t, e_{t-1}) = 0$. The alternative hypothesis is that autocorrelation is present. A test statistic close to 2 indicates no autocorrelation, values substantially below 2 suggest positive autocorrelation, and values above 2 suggest negative autocorrelation.

A way to deal with AR(1)-autocorrelation is to use the Cochrane-Orcutt technique, which transforms the regression model to eliminate the autocorrelation structure in the residuals [24].

Suppose the error term follows a first-order autoregressive process:

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t, \quad |\rho| < 1$$

where $u_t \sim$ i.i.d. $N(0, \sigma^2)$. The condition $|\rho| < 1$ ensures stationarity of the error process and is typically satisfied after proper normalization.

The autoregressive coefficient ρ is estimated by regressing the residuals from the original OLS model on their lagged values. Using the estimated $\hat{\rho}$, the original regression model:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \cdots + \beta_k x_{k,t} + \varepsilon_t$$

is transformed using:

$$y_t^* = y_t - \rho y_{t-1}, \quad x_{j,t}^* = x_{j,t} - \rho x_{j,t-1}$$

OLS is then applied to the transformed model $y_t^* = \beta_0(1-\rho) + \beta_1 x_{1,t}^* + \cdots + \beta_k x_{k,t}^* + u_t$, yielding estimates corrected for autocorrelation. While ρ appears in both the transformation and the intercept term, it is treated as fixed once estimated, and is not re-estimated within the final regression step.

The Breusch–Godfrey (BG) test [21] detects autocorrelation in the residuals of a regression model, including higher-order autocorrelation. The auxiliary regression is:

$$e_t = \alpha + \sum_{j=1}^{p} \beta_j X_{j,t} + \sum_{k=1}^{m} \rho_k e_{t-k} + u_t, \tag{2.25}$$

where e_t are the residuals from the original model, $X_{j,t}$ are the predictors, e_{t-k} are lagged residuals (up to order m), and u_t is a new error term. The null hypothesis is that there is no autocorrelation up to lag m, i.e., $\rho_1 = \cdots = \rho_m = 0$. The test statistic is:

$$LM = nR^2 \sim \chi_m^2,$$

where R^2 comes from the auxiliary regression, χ_m^2 is the chi-squared distribution with p degrees of freedom, and m is the number of lags tested. A significant result suggests autocorrelated residuals.

3 Method

3.1 Data Preparation, Segmentation, and Feature Engineering

The dataset consists of Norion Bank's entire portfolio of Swedish consumer credit loans spanning from June 2019 to February 2024, (excluding August 2023), covering a total of 56 months. The dataset includes all contracts that were active at any month during this period. With a monthly average of above 50000.
For every month t, each contract l is assigned a probability of default over the coming 12 months, $\mathrm{PD}_{l,t}^{\mathrm{12m,Norion}}$ (see Equation (2.1)). These values are generated by Norion Bank's internally developed credit risk models, which rely on borrower characteristics, historical performance, and other contract-specific information. In this study, $\mathrm{PD}_{l,t}^{\mathrm{12m,Norion}}$, is provided directly by the bank as part of the dataset and is treated as a fixed input. At each month t,and contract l, the dataset also includes whether the contract defaulted within the coming 12 months $(t + 12)$, or not.

Total portfolio

Our first model is what we call the total portfolio ($\mathcal{M}^{total}$). Here we use all the contracts active at each month and aggregate them to the portfolio-level as described in Section 2.2. From this we compute the spread s_t, as the difference between observed default rates and the predicted default rate. In Table 3.1 we can see statistics on the spread, $\mathrm{s}_t^{\mathrm{Total}}$, of $\mathcal{M}^{total}$. In Figure 3.1, we can see that the current $\mathrm{PD}^{\mathrm{12m,\ Norion}}$-model underestimates risk in most of its monthly predictions - the number of defaults is higher than the predicted defaults.

Table 3.1: Summary of the observed spread, $\mathrm{s}_t^{\mathrm{Total}}$, for the total portfolio, $\mathcal{M}^{\mathrm{Total}}$.

		Spread		
Portfolio	**Avg. #Contracts/Month**	**Min**	**Max**	**Mean**
Total Portfolio	>50 000	−0.0057	0.0274	0.0120

Next, the portfolio is segmented to assess model performance across different borrower types, and to investigate which macroeconomic indicators are most important for these sub-groups. Apart from the total portfolio ($\mathcal{M}^{Total}$), the portfolio is segmented based on three key borrower characteristics;

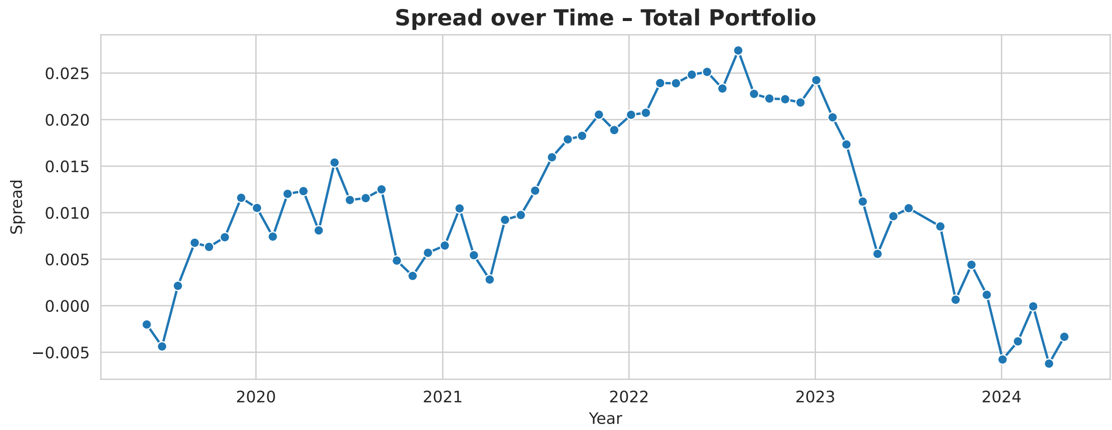

Figure 3.1: Spread s_t^{Total} over time for the total portfolio, $\mathcal{M}^{\text{Total}}$.The figure shows that the current $\text{PD}^{\text{12m, Norion}}$-model generally underestimates the risk. Clear peaks at rough macroeconomic times like COVID-19.

- Age Group, $\mathcal{M}^{(\text{Age},\cdot)}$
- Mortgage Holder status, $\mathcal{M}^{(\text{MH},\cdot)}$
- Co-Borrower status, $\mathcal{M}^{(\text{CB},\cdot)}$

For each segmentation, the aggregation procedures described in Section 2.2 is repeated within each subgroup, yielding segment-specific time series for $\overline{\text{PD}}_t^{12m,(\cdot,\cdot)}$, $\overline{\text{OD}}_t^{12m,(\cdot,\cdot)}$, and the corresponding spreads $s_t^{(\cdot,\cdot)}$, allowing us to analyze each subgroup's performance separately.

Age Group Segmentation $\mathcal{M}^{(\text{Age},\cdot)}$

In the age segmentation, contracts are grouped into three categories based on the age of the borrower:

- Young: 18–35 years, $\mathcal{M}^{(\text{Age},\text{Young})}$
- Middle: 36–55 years, $\mathcal{M}^{(\text{Age},\text{Middle})}$
- Older: 56 years and above, $\mathcal{M}^{(\text{Age},\text{Older})}$

In Table 3.2 we see average contract count for each month, and information about the spread, $s_t^{(\text{Age},\cdot)}$, for $\mathcal{M}^{(\text{Age},\cdot)}$. Figure 3.2 shows the spread, $s_t^{(Age,\cdot)}$ for each month.

Table 3.2: Summary of the spread, $s_t^{(\text{Age},\cdot)}$, for each age sub-group

		Spread		
Age Group	**Share of contracts**	**Min**	**Max**	**Mean**
Young	16.5 %	−0.0123	0.0502	0.0237
Middle	48.2 %	−0.0127	0.0256	0.0078
Older	35.3 %	−0.0036	0.0200	0.0098

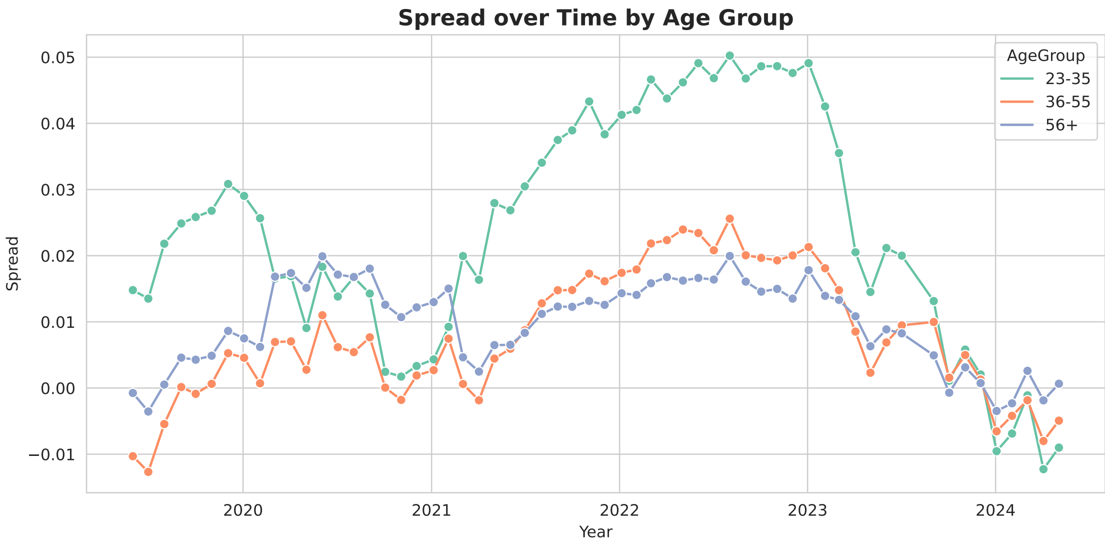

Figure 3.2: Spread, $s_t^{(\text{Age},\cdot)}$, over time by age group. We see that the current model underestimates the spread. Especially, $s_t^{(\text{Age},\text{Young})}$, for the young group $\mathcal{M}^{(\text{Age},\text{Young})}$.

Mortgage Holder Segmentation, $\mathcal{M}^{(\text{MH},\cdot)}$

For the mortgage segmentation, contracts are segmented based on whether the borrower holds a mortgage. In this case we mean if they have apartment or house loans:

- Mortgage holder, $\mathcal{M}^{(\text{MH},\text{Yes})}$
- No mortgage, $\mathcal{M}^{(\text{MH},\text{No})}$
- Unknown status, $\mathcal{M}^{(\text{MH},\text{Unknown})}$

In Table 3.3 we can see monthly average contract count, and information about the spread, $s_t^{(\text{MH},\cdot)}$, for $\mathcal{M}^{(\text{MH},\cdot)}$. The spread, $s_t^{(\text{MH},\cdot)}$, for each month is shown in Figure 3.3 .

Table 3.3: Summary of the spread, $s_t^{(\text{MH},\cdot)}$, for each mortgage holder sub-group.

		Spread		
Mortgage Status	**Share of Contracts**	**Min**	**Max**	**Mean**
Yes	13.3 %	−0.0360	0.0029	−0.0083
No	27.5 %	−0.0019	0.0459	0.0280
Unknown	59.2 %	−0.0047	0.0219	0.0092

Co-Borrower Segmentation, $\mathcal{M}^{(\text{CB},\cdot)}$

The contracts are also segmented based on their co-borrower status. Meaning, if there are more than one individual connected to the credit, then its classified as having a co-borrower. The segmentation is:

- With co-borrower, $\mathcal{M}^{(\text{CB},\text{Yes})}$

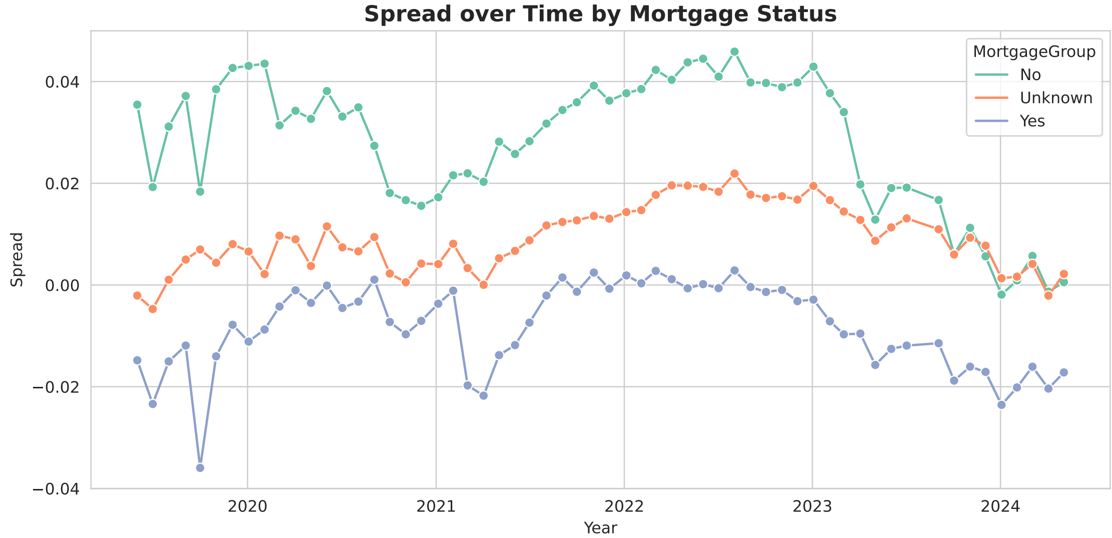

Figure 3.3: Spread over time by mortgage status,$s_t^{(MH,\cdot)}$. We can see a clear separation of the different segmentation. Persistent underestimation of risk for the non-mortgage group, and over-estimation for the Yes group.

- Without co-borrower, $\mathcal{M}^{(CB,No)}$

Table 3.4 shows the monthly average contract count, and information about the spread, $s_t^{(CB,\cdot)}$, for $\mathcal{M}^{(CB,\cdot)}$. A time-series plot of the spread, s_t is displayed in Figure 3.4.

Table 3.4: Summary of the spread, $s_t^{(CB,\cdot)}$, by Co-Borrower status.

		Spread		
Co-Borrower Status	**Share of Contracts**	**Min**	**Max**	**Mean**
No	91.7 %	−0.0064	0.0312	0.0137
Yes	8.3 %	−0.0429	−0.0039	−0.0167

Macroeconomic Data

The macroeconomic indicators of interest from Section 2.3 are collected from January 2016 to the latest available month. While most data is reported monthly, certain indicators vary in reporting frequency. Specifically, GDP growth is reported quarterly. To match with the other indicators (and credit data) monthly frequency, we use linear interpolation to convert quarterly observations into monthly estimates. Conversely, the KIX index is reported daily, and monthly averages are calculated to produce monthly observations. A detailed description of all macroeconomic indicators is presented in Appendix A.

The raw macroeconomic variables X_t are then transformed using the feature engineering from Subsection 2.4.1. Specifically, each macroeconomic variable is trans-

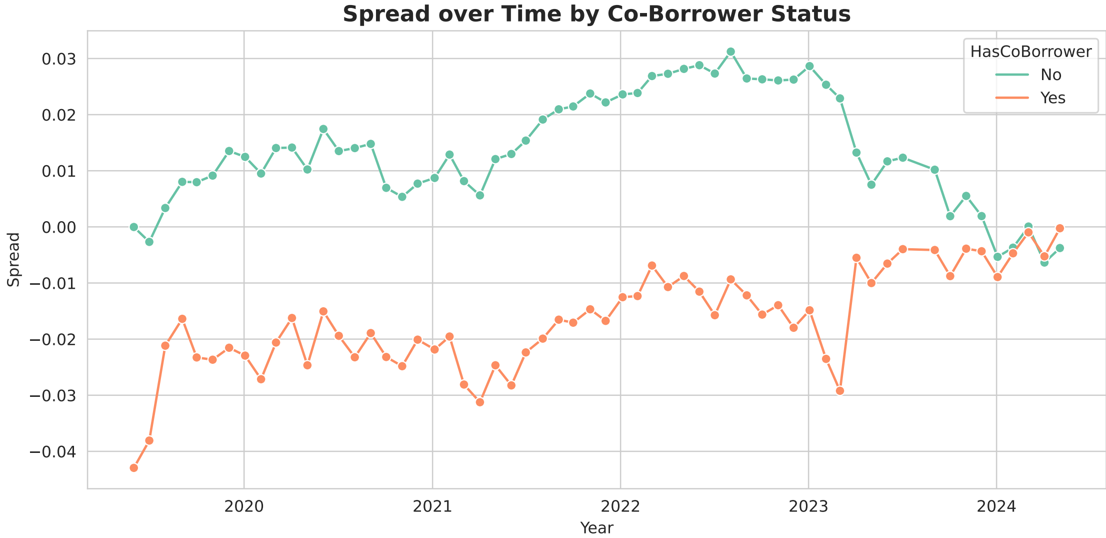

Figure 3.4: Spread,$s_t^{(CB,\cdot)}$, over time by co-borrower status $\mathcal{M}^{(CB,\cdot)}$. Contracts with Co-Borrowers is almost always overestimated in risk, while the reverse is true about single-borrowers.

formed into: Lagged features (2.6), delta features (2.8), and rolling average feature (2.7). The use of these transformations is justified based on the previously discussed lagged effects of macroeconomy on credit risk.

For the total portfolio $\mathcal{M}^{\text{Total}}$, and each segmented portfolio $\mathcal{M}^{(\cdot,\cdot)}$, the transformed macroeconomic features are combined with the credit portfolio data. Then the dataset is divided into training and testing subsets based on a predefined cutoff date, February 4, 2022. Observations occurring on or before this date form the training set, while those after the split date comprise the test set. The split-date was chosen so that we have 12-months of unseen data at the end in the test-data, while the first 12 months has the issue of overlapping period, further discussed in Subsection 2.2.1.

3.1.1 Feature Selection

Feature selection is used to identify the most relevant macroeconomic indicators affecting the spread, and to address high dimensionality, multicollinearity, and interpretability. A three-step selection process is used: (1) correlation filtering to retain features strongly linked to the spread, (2) ElasticNet regularization to reduce redundancy, and (3) iterative VIF filtering to remove multicollinear features. The procedure below is done for each portfolio separately.

Initially, Pearson correlations (Equation (2.10)) are computed between the transformed macroeconomic indicators and each separate model spread ($s_t^{(\cdot,\cdot)}$, and s_t^{Total}). To deal with the large selection of candidates, automated selection is done using the following rules:

- Lag features (Equation 2.6): Retain lagged features with absolute correlation ≥ 0.6. To avoid redundancy, we only select one positively correlated and one negatively correlated lag per macroeconomic variable. Chosen lags should represent either clear peaks or 'stable' correlation patterns over several consecutive lags.
- Rolling features (Equation 2.7): Only retain the rolling feature with the highest absolute correlation (if ≥ 0.6).
- Delta features (Equation 2.8): Again, select only the feature with highest absolute correlation (if ≥ 0.6). Additionally, a second delta feature can be chosen if it has sufficiently different transformation window.

This is done for each separate model so that each model has their own set of correlated features.

Following the initial correlation-based selection, the retained macroeconomic features still contain redundant and correlated information. To reduce the dimensionality further, and mitigate multicollinearity, ElastiNet regularization (see Section 2.5, and Equation (2.11)) is applied. ElasticNet´s hyperparameters are calibrated via 5-fold cross-validation (discussed in Section 2.5), selecting the values that minimizes the average predictive error from Equation (2.12). Note that this 5-fold cross-validation is only done on the train set. The test set is untouched and only used to asses out-of-sample performance. Features whose coefficients are shrunk to zero are discarded.

The remaining features from the ElasticNet stage may still exhibit problematic multicollinearity, which mayaffect the stability and interpretability of the OLS regression negatively. Therefore we use an iterative VIF filtering approach (see Subsection 2.5). Specifically, we calculate the VIF (Equation (2.13)) for each feature, then iteratively remove the feature with highest VIF value-provided it exceeds the threshold of 10. This process is repeated until all remaining features have acceptable VIF values below the threshold.

As a last step in the feature selection process, Norion´s own $\text{PD}_t^{\text{12m,Norion}}$-model variable is added as a feature (denoted MeanPredictionPd12m), regardless of it's VIF. This inclusion is motivated by its baseline predictive ability and its role in anchoring the prediction toward a reasonable baseline.

3.1.2 Linear Regression and AR(1) Correction

Using the features retained from the above feature selection process, the final model is first estimated using standard Ordinary Least Squares (OLS) regression, applied to the selected predictor variables standardized to zero mean and unit variance. Standardization improves coefficient stability and enables meaningful comparison across predictors.

To detect autocorrelation, the Durbin–Watson (DW) statistic is computed on the residuals (Equation 2.24). If DW < 1.5, indicating first-order autocorrelation, the model is corrected using the Cochrane–Orcutt method. The AR(1)-corrected model is only retained if it improves train-set performance, measured by Mean Absolute Error (MAE).

A final check of the correlation between the final features is done using correlation heatmaps. Manual inspection removes highly correlated features and the model is then refitted. If the model MAE suffers too much when removing features - and the correlation is manageable - then, the feature is kept.

3.1.3 Model Evaluation and Validation

In addition to MAE, residual diagnostics are performed to validate the model assumptions, including tests for heteroskedasticity (Breusch–Pagan test), autocorrelation (Durbin–Watson statistic), and normality (Shapiro–Wilk test). These diagnostics help ensure that the linear regression assumptions hold and that the model's inferences are reliable. The model is evaluated both on the training set and on the unseen test set to assess its generalization performance.

3.2 Limitations and Data Reliability

While the macroeconomic indicators originate from authoritative and reputable institutions, some limitations should be acknowledged:

- Data revisions: Indicators, particularly GDP and inflation, are frequently revised after their initial publication as more comprehensive data become available.
- Interpolation uncertainty: Monthly interpolation of GDP, and month-averaging of KIX index introduces estimation uncertainty, as interpolated values may not accurately reflect actual economic conditions within each month.
- A particular challenge arises from differences in the timing of macroeconomic data publication and the timing of credit risk valuation at Norion Bank. Typically, macroeconomic indicators are published at month-end, while IFRS 9 risk modeling at the bank is performed at the beginning of the month. To address this timing mismatch, end-of-month macroeconomic reporting dates are systematically ("moved forward") replaced with IFRS 9 portfolio evaluation dates. This ensures that the temporal alignment between macroeconomic trends and portfolio risk assessment remains accurate and relevant. Given the assumption that the influence of macroeconomic indicators on credit risk is inherently lagged, the impact of data revisions and interpolation errors on modeling outcomes is considered acceptable within the context of this analysis..
- Period overlapping: Given the 12-month horizon used to define realized defaults, a problem arises during the initial 12 months of the test set, where training and test periods partially overlap due to the lagged nature of default outcomes. This overlap can inflate performance metrics, thereby overstating the true predictive capability of the model in an out-of-sample context.

4

Results

In this chapter we present empirical findings from the modeling. Plots, tables and diagnostics are used to describe the results. The chapter is divided so that each portfolio and borrower segmentation is presented in a clear manner.

4.1 Total Portfolio, $\mathcal{M}^{\textbf{total}}$

In the initial correlation-based feature selection, 45 features were retained. After ElasticNet regularization, 13 features remained (see Table 4.1), and VIF filtering removed seven features due to high VIF (> 10). Additionally, two features were removed after correlation heatmap inspection. Correlation heatmap of final features can be seen in Figure 4.2.

The Durbin–Watson statistic for training residuals indicated first-order autocorrelation. Consequently, AR(1) correction via the Cochrane–Orcutt method was applied. The AR(1)-corrected model yielded the better train MAE, and is therefore picked as final. Table 4.3 show the coefficients of the final model, and Table 4.4 show some diagnostics on the residuals. Residuals do not appear to be normally distributed. Residuals have constant variance, and the Breusch–Godfrey test shows no strong evidence of higher-order autocorrelation.

The final AR(1)-corrected model has an R^2 of 0.522 on the test set and achieves a test MAE of 0.0055 (see Table 3.1). Illustration of the observed spread, $\mathrm{s}_t^{\mathrm{Total}}$, and the predicted spread $\hat{s}_t^{\mathrm{Total}}$, as well as the residual,e_t^{Total}, is shown in Figure 4.1.

4.2 Age Group Segmentations, $\mathcal{M}^{(\textbf{Age},\cdot)}$

Separate models were estimated for the three age groups. Feature selection, VIF filtering, and model-choice criteria use the same procedure as for the total portfolio above. Table 4.5 lists the final (standardised) coefficients for all age sub-groups.

Age: Young (18–35), $\mathcal{M}^{(\textbf{Age},\textbf{Young})}$

Initially, 43 macroeconomic features were identified based on correlation criteria, 7 remained after ElasticNet (see Table B.1), VIF filtering eliminated 5 additional features. Correlation heatmap of final feature can be seen in Figure B.1, and coefficients are shown in Table 4.5. The DW-statistic indicated first-order autocorrelation. The

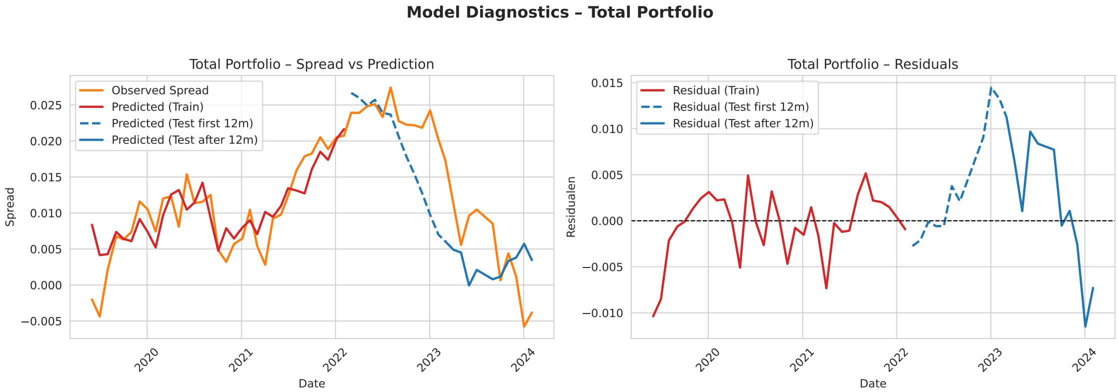

Figure 4.1: The model's predicted spread for the total aggregated portfolio, $\mathcal{M}^{\text{Total}}$. The left panel shows the observed spread, s_t^{Total}, and the predicted spread, $\hat{s}_t^{\text{Total}}$, over time. the plot illustrates that the prediction mostly follows the pattern of the observed value. The right panel displays the residuals, e_t^{Total}. Residuals remain low even on the test set. Note the different scales on the y-axes.

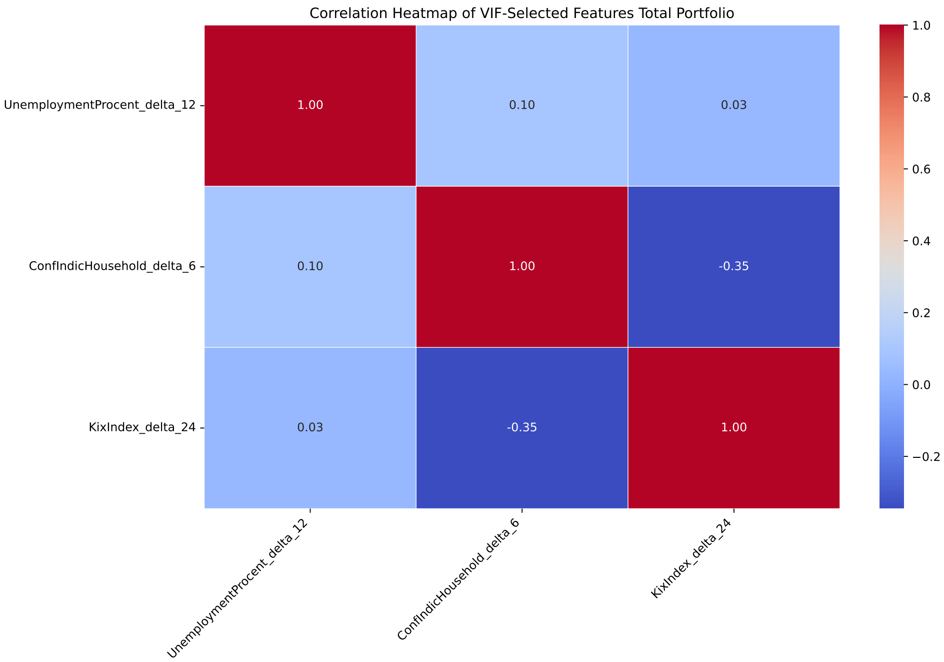

Figure 4.2: Heatmap of the correlation of the final features for the total portfolio, $\mathcal{M}^{\text{Total}}$.

Table 4.1: Features selected by ElasticNet on the $\mathcal{M}^{\text{Total}}$ portfolio. Features are ordered by the magnitude of their coefficients.

Feature	**ElasticNet Coefficient**	**Sign**
MeanPredictionPd12m	-6.49×10^{-3}	−
ConfIndicHousehold_delta_6	-1.12×10^{-3}	−
ServiceSectorIndic_lag_7	8.99×10^{-4}	+
RepoRate_delta_12	8.24×10^{-4}	+
KPIF_lag_15	7.80×10^{-4}	+
GDPChange_roll_12	5.64×10^{-4}	+
ConfIndicHousehold_lag_7	4.92×10^{-4}	+
RetailTradeIndic_lag_12	3.63×10^{-4}	+
KixIndex_delta_24	1.81×10^{-4}	+
GDPChange_lag_26	1.30×10^{-4}	+
KpifFixedBaseLine_delta_3	-2.36×10^{-4}	−
UnemploymentProcent_delta_12	-2.36×10^{-4}	−
KpifFixedBaseLine_delta_12_24	8.9×10^{-5}	+

Table 4.2: Model performance of $\mathcal{M}^{\text{Total}}$ for the OLS and the AR(1)-corrected model. The latter chosen as final model based on better train MAE.

Metric	**Train**		**Test**	
	OLS	AR(1)	OLS	AR(1)
MAE	0.00260	0.00255	0.00640	0.00554
R^2	0.694	0.675	0.409	0.522
Adjusted R^2	0.662	0.641	0.315	0.446

Table 4.3: Coefficients from the final AR(1)-corrected model (standardized variables) for $\mathcal{M}^{\text{Total}}$, using the Cochrane–Orcutt method. Features are ordered by the magnitude of their coefficients. t-statistics and p-values are calculated on the transformed model.

Feature	**Coefficient**	**t-statistic**	**p-value**
KixIndex_delta_24	−0.0034	−4.30	0.0002
ConfIndicHousehold_delta_6	−0.0032	−4.62	< 0.001
UnemploymentProcent_delta_12	−0.0020	−3.16	0.0038
Intercept	0.0074	13.87	< 0.001

Table 4.4: Summary of residual diagnostic tests (retained AR(1)-model) on $\mathcal{M}^{\text{Total}}$.

Test	Statistic	p-value
Shapiro–Wilk (normality)	0.9171	0.0152
Breusch–Pagan (LM test)	1.8899	0.5956
Breusch–Pagan (F test)	0.5872	0.6283
Breusch–Godfrey (LM test)	3.2712	0.1948
Breusch–Godfrey (F test)	1.4855	0.2443

AR(1)-corrected model exhibited worse train MAE. Consequently, the OLS model was retained as the final choice. Model performance metrics for both OLS and AR(1)-corrected models are shown in Table B.2.

Diagnostic tests on the residuals of the final OLS model are summarized in Table B.3. Residuals indicate deviations from normality, and indication of higher-order autocorrelation. The final OLS model has a test R^2 of 0.804, and a test MAE of 0.00683. Figure 4.3 shows plots of the observed spread $\mathrm{s}_t^{(\text{Age,Young})}$ and the predicted spread, $\hat{s}_t^{(\text{Age,Young})}$, as well as the residual, $e_t^{(\text{Age,Young})}$.

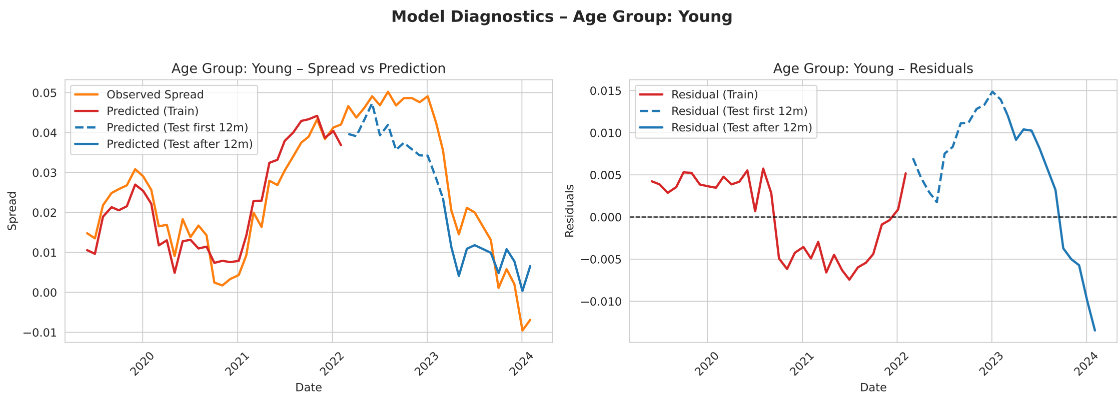

Figure 4.3: The model's predicted spread for the age group 23–35, $\mathcal{M}^{(\text{Age,Young})}$. The left panel shows the observed spread, $s_t^{(\text{Age,Young})}$, and the predicted spread, $\hat{s}_t^{(\text{Age,Young})}$, over time. Prediction follows the observed value well. The right panel displays the residuals, $e_t^{(\text{Age,Young})}$. Residuals are somewhat larger on the test set. Note the different scales on the y-axes.

Age: Middle (36–55), $\mathcal{M}^{(\text{Age,Middle})}$

The initial correlation-based selection retained 19 features, ElasticNet reduced this number to 14 features (see Table B.4), VIF filtering removed 11 features. Additionally one feature was manually removed for high correlation, see heatmap in Figure B.2. Model performance is summarized in Table B.5. The DW-statistic indicated autocorrelation. The AR(1) correction did not improve train performance.

Table 4.5: Final model coefficients by age group, $\mathcal{M}^{(\text{Age},\cdot)}$ (standardised). Features are ordered by the magnitude of their coefficients (except Intercept).

Age Group	Feature	Coef.	t	p
Young	MeanPredictionPd12m	−0.0085	−10.27	< 0.001
	Consumption_delta_24	0.0041	4.95	< 0.001
	Intercept	0.0223	31.14	< 0.001
Middle	GDPChange_lag_30	0.0040	4.29	< 0.001
	Consumption_delta_12_24	−0.0030	−3.22	0.003
	Intercept	0.0051	6.34	< 0.001
Older	MeanPredictionPd12m	−0.0029	−3.22	0.003
	GDPChange_lag_30	0.0021	2.40	0.023
	Intercept	0.0102	15.82	< 0.001

The OLS- model was selected, coefficients are shown in Table 4.5. Diagnostic results (Table B.6) show that residuals deviate from normality. Heteroskedasticity and higher-order autocorrelation are detected. In Figure 4.4 plots of the observed spread $\mathrm{s}_t^{(\text{Age,Middle})}$ and the predicted spread, $\hat{s}_t^{(\text{Age,Middle})}$, as well as the residual, $e_t^{(\text{Age,Middle})}$ is shown.

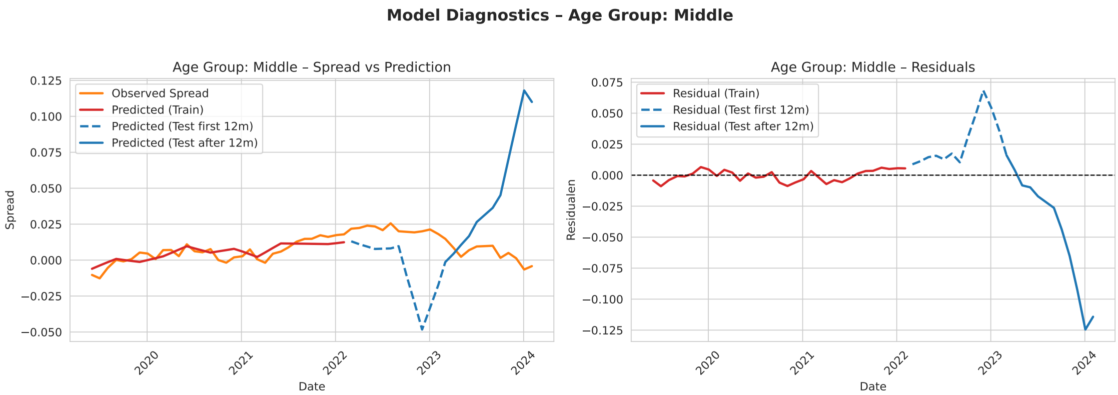

Figure 4.4: The model's predicted spread for the age group 36–55, $\mathcal{M}^{(\text{Age,Middle})}$. The left panel shows the observed spread, $s_t^{(\text{Age,Middle})}$, and the predicted spread, $\hat{s}_t^{(\text{Age,Middle})}$, over time. The prediction performs well on train data but fails completely on test data. The right panel displays the residuals, $e_t^{(\text{Age,Middle})}$. Note the different scales on the y-axes.

Age: Older (56+), $\mathcal{M}^{(\textbf{Age,Older})}$

The initial correlation-based selection retained 5 features. ElasticNet kept all (see Table B.7). VIF filtering removed two features. One additional was removed manually due to high correlation with other features. Final correlation heatmap can be seen in Figure B.3. Model performance (Table B.8) indicates autocorrelation.

Applying AR(1) correction did not improve train MAE. The OLS-model was therefore selected. The final model has train R^2 of 0.631. Negative test R^2 shows bad generalization. The OLS-model coefficients can be seen in Table 4.5. Diagnostics (Table B.9) revealed that residuals deviate strongly from normality, exhibit heteroscedasticity. In Figure 4.5 plots of the observed spread $\mathrm{s}_t^{(\mathrm{Age,Older})}$ and the predicted spread, $\hat{s}_t^{(\mathrm{Age,Older})}$, as well as the residual, $e_t^{(\mathrm{Age,Older})}$ is shown.

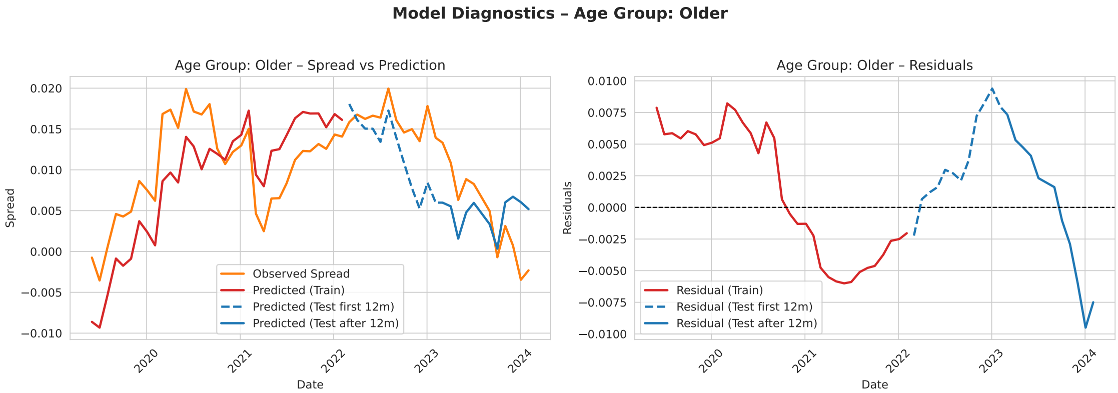

Figure 4.5: The model's predicted spread for the age group 56+, $\mathcal{M}^{(\mathrm{Age,Older})}$. The left panel shows the observed spread, $s_t^{(\mathrm{Age,Older})}$, and the predicted spread, $\hat{s}_t^{(\mathrm{Age,Older})}$, over time. Predictions follow the pattern of observed alright, but deviates slightly in the final months. The right panel displays the residuals, $e_t^{(\mathrm{Age,Older})}$. A bit larger residual errors on the test set compared to train. Note the different scales on the y-axes.

4.3 Mortgage Holder Segmentation, $\mathcal{M}^{(\mathbf{MH},\cdot)}$

Mortgage Segment: Yes, $\mathcal{M}^{(\mathbf{MH,Yes})}$

Correlation-based feature selection initially identified four features. ElasticNet kept all (see Table B.10). Three features remained after VIF filtering. Model performance is summarized in Table B.11. Correlation heatmap is shown in Figure B.4. The AR(1)-corrected model did not improve train MAE. The model achieves an R^2 of 0.811 on the training set. However, the test R^2 is highly negative. Final model coefficients are shown in Table 4.6. Residual diagnostics (Table B.12) indicate non-normality, heteroscedasticity, and higher-order autocorrelation. Figure 4.6 shows plots of the observed spread $\mathrm{s}_t^{(\mathrm{MH,Yes})}$ and the predicted spread, $\hat{s}_t^{(\mathrm{MH,Yes})}$, as well as the residual, $e_t^{(\mathrm{MH,Yes})}$.

Mortgage Segment: No,$\mathcal{M}^{(\mathbf{MH,No})}$

The initial correlation-based selection identified 14 features. ElasticNet reduced this to six (see Table B.13). After VIF filtering, three features remained. Model performance is summarized in Table B.14, and correlation heatmap in Figure B.5.

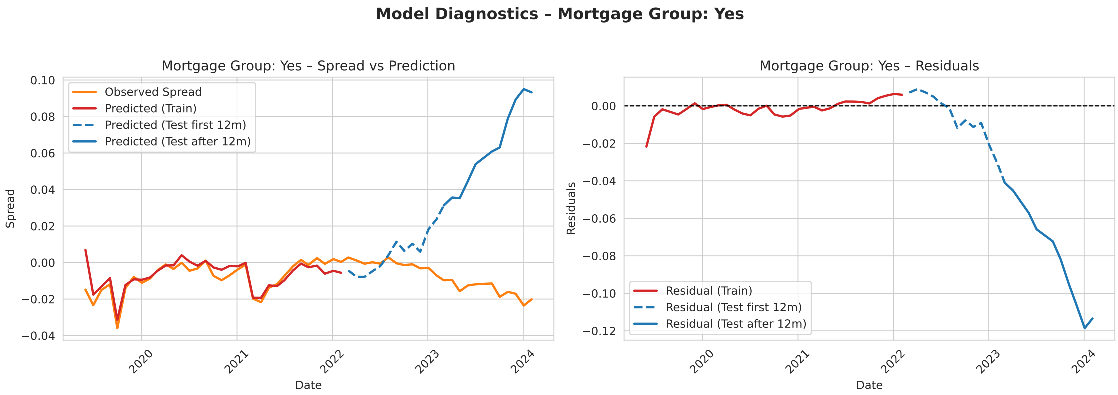

Figure 4.6: The model's predicted spread for the mortgage segment with a mortgage, $\mathcal{M}^{(\text{MH,Yes})}$. The left panel shows the observed spread, $s_t^{(\text{MH,Yes})}$, and the predicted spread, $\hat{s}_t^{(\text{MH,Yes})}$, over time. The model performs poorly on the test data. The right panel displays the residuals, $e_t^{(\text{MH,Yes})}$, across the same period. Note the different scales on the y-axes.

Table 4.6: Final model coefficients by mortgage group $\mathcal{M}^{(\text{MH},\cdot)}$ (standardised). Features are ordered on the magnitude of their coefficients.

Mortgage Group	Feature	Coef.	t	p
Yes	MeanPredictionPd12m	−0.0049	−7.04	< 0.001
	RepoRate_roll_3	0.0033	3.77	0.001
	GDPChange_lag_30	0.0031	3.52	0.001
	Intercept	−0.0079	−11.49	< 0.001
No	MeanPredictionPd12m	−0.0032	−2.83	0.008
	ConfIndicHousehold_delta_6	−0.0029	−2.57	0.016
	GDPChange_roll_12	0.0024	1.78	0.085
	Intercept	0.0306	33.52	< 0.001
Unknown	KixIndex_delta_12_24	−0.0036	−4.32	< 0.001
	Intercept	0.0039	7.69	< 0.001

Although Durbin–Watson indicated autocorrelation, AR(1) correction did not improve train MAE. The OLS model was retained. The final model achieves an R^2 of 0.654 on the training set and 0.700 on the test set. Final model coefficients are listed in Table 4.7. Residual diagnostics (Table B.15) show residuals appear normally distributed, with no sign of heteroscedasticity. However, Breusch–Godfrey indicates some higher-order autocorrelation. In Figure 4.7 plots of the observed spread $\mathrm{s}_t^{(\mathrm{MH,No})}$ and the predicted spread, $\hat{s}_t^{(\mathrm{MH,No})}$, as well as the residual, $e_t^{(\mathrm{MH,No})}$ are shown.

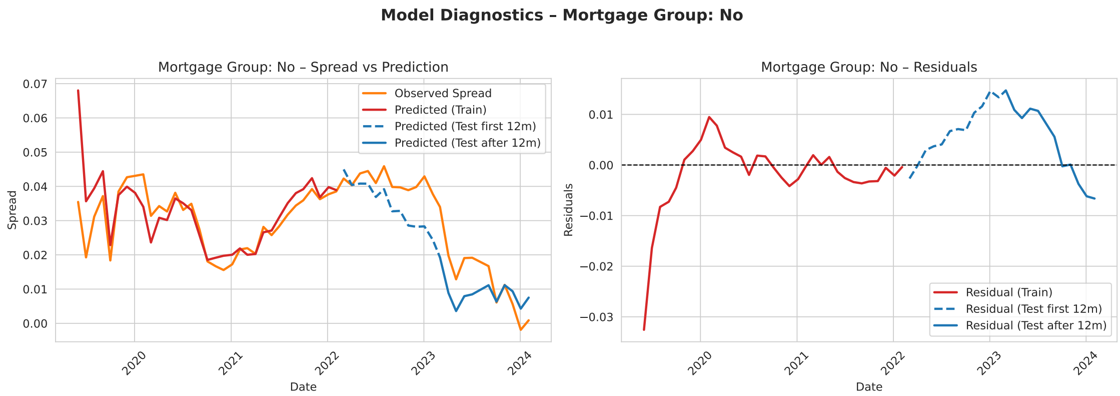

Figure 4.7: The model's predicted spread for the mortgage segment without a mortgage, $\mathcal{M}^{(\mathrm{MH,No})}$. The left panel shows the observed spread, $s_t^{(\mathrm{MH,No})}$, and the predicted spread, $\hat{s}_t^{(\mathrm{MH,No})}$, over time. Prediction closely follows the pattern of observed value well. The right panel displays the residuals, $e_t^{(\mathrm{MH,No})}$. A bit bigger residuals on the test set, than on train. Note the different scales on the y-axes.

Mortgage Segment: Unknown, $\mathcal{M}^{(\mathrm{MH,Unknown})}$

The initial correlation-based selection retained 17 features. ElasticNet reduced this to 9 features (see Table B.16). Two features remained after VIF. Still, there was high correlation between features, resulting in picking the one with highest train MAE when fitting a solo-model. Model performance is summarized in Table B.17. The DW-statistic indicated autocorrelation. AR(1) correction marginally improved train MAE from 0.00264 to 0.00263, so the AR(1)-corrected model was selected. R^2 on the training set was 0.508, test R^2 was negative (-2.99). Final model coefficients are listed in Table 4.6. Table B.18 indicate that residuals are approximately normally distributed and no heteroscedasticity. However, Breusch–Godfrey shows higher-order autocorrelation. We can see plots of the observed spread $\mathrm{s}_t^{(\mathrm{MH,Unknown})}$ and the predicted spread, $\hat{s}_t^{(\mathrm{MH,Unknown})}$, as well as the residual, $e_t^{(\mathrm{MH,Unknown})}$ in Figure 4.8.

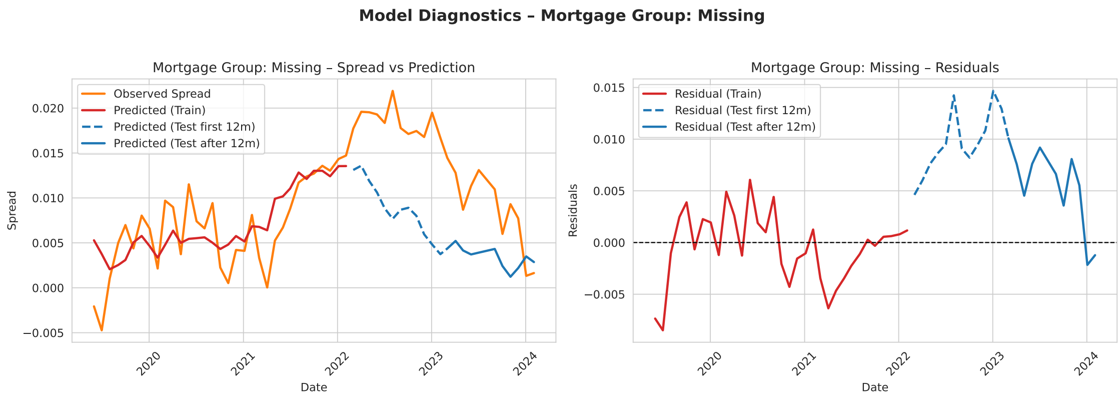

Figure 4.8: The model's predicted spread for the Unknown-mortgage status, $\mathcal{M}^{(\mathrm{MH,Unknown})}$. The left panel shows the observed spread, $s_t^{(\mathrm{MH,Unknown})}$, and the predicted spread, $\hat{s}_t^{(\mathrm{MH,Unknown})}$, over time. Prediction is a bit off, but still captures the overall movement of the observation. The right panel displays the residuals, $e_t^{(\mathrm{MH,Unknown})}$. Residuals are biased towards positive. Note the different scales on the y-axes.

4.4 Co-Borrower Segmentation, $\mathcal{M}^{(\mathrm{CB},\cdot)}$

Co-Borrower Segment: Yes, $\mathcal{M}^{(\mathrm{CB,Yes})}$

The initial correlation-based selection retained three features. ElasticNet regularization retained two, see Table B.19. After VIF filtering, only one feature remained. The DW-statistic indicated autocorrelation, AR(1) correction decreased train MAE from 0.00366 to 0.00355. The AR(1)-corrected model was retained. Coefficients from final model are listed in Table 4.7. Model performance is summarized in Table B.20. Residual diagnostics (Table B.21) show that residuals are approximately normally distributed, with no evidence of heteroscedasticity. In Figure 4.9 plots of the observed spread $s_t^{(\mathrm{CB,Yes})}$ and the predicted spread, $\hat{s}_t^{(\mathrm{CB,Yes})}$, as well as the residual, $e_t^{(\mathrm{CB,Yes})}$ is shown.

Table 4.7: Final model coefficients by co-borrower group $\mathcal{M}^{(\mathrm{CB},\cdot)}$ (standardised).

Co-Borrower Group	Feature	Coef.	t	p
Yes	GDPChange_lag_30	0.0036	3.55	0.001
	Intercept	−0.0146	−20.97	< 0.001
No	GDPChange_lag_26	0.0050	6.42	< 0.001
	Intercept	0.0124	16.00	< 0.001

Co-Borrower Segment: No, $\mathcal{M}^{(\mathrm{CB,No})}$

The initial correlation-based selection retained 20 features, with ElasticNet reducing this to 10 (see Table B.22). After VIF filtering, seven variables were removed. Three

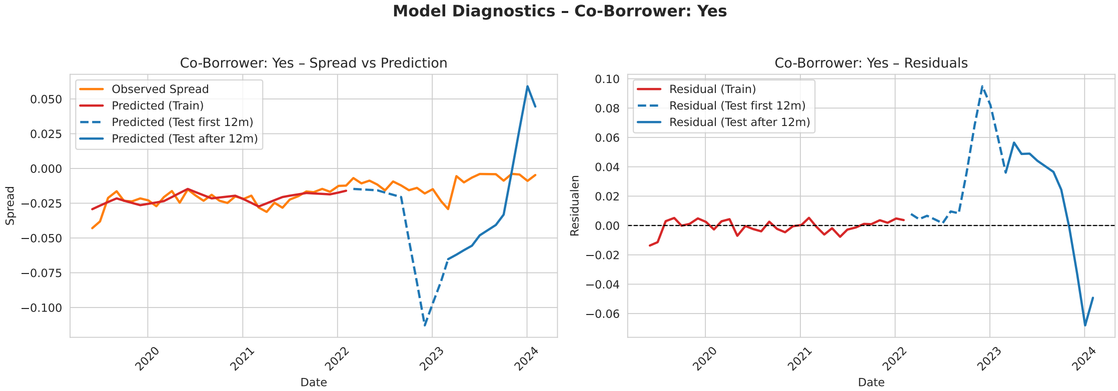

Figure 4.9: The model's predicted spread for the co-borrower segment with a co-borrower, $\mathcal{M}^{(\mathrm{CB,Yes})}$. The left panel shows the observed spread, $s_t^{(\mathrm{CB,Yes})}$, and the predicted spread, $\hat{s}_t^{(\mathrm{CB,Yes})}$, over time. The model performs poorly on the test set. The right panel displays the residuals, $e_t^{(\mathrm{CB,Yes})}$. Note the different scales on the y-axes.

additional features were removed due to high intercorrelation with each other (final correlation heatmap in Figure B.8. Model performance is summarized in Table B.23. Although autocorrelation was detected, the AR(1) correction increased train MAE from 0.00337 to 0.00359. The OLS model was retained. It achieved an R^2 of 0.570 on the training set and -41.242 on the test set. Final model coefficients are shown in Table 4.7. Residual diagnostics (Table B.24) indicate that residuals are normally distributed, with no evidence of heteroscedasticity. However, the Breusch–Godfrey test indicated significant higher-order autocorrelation. We can see plots of the observed spread $\mathrm{s}_t^{(\mathrm{CB,No})}$ and the predicted spread, $\hat{s}_t^{(\mathrm{CB,No})}$, as well as the residual, $e_t^{(\mathrm{CB,No})}$ in Figure 4.10.

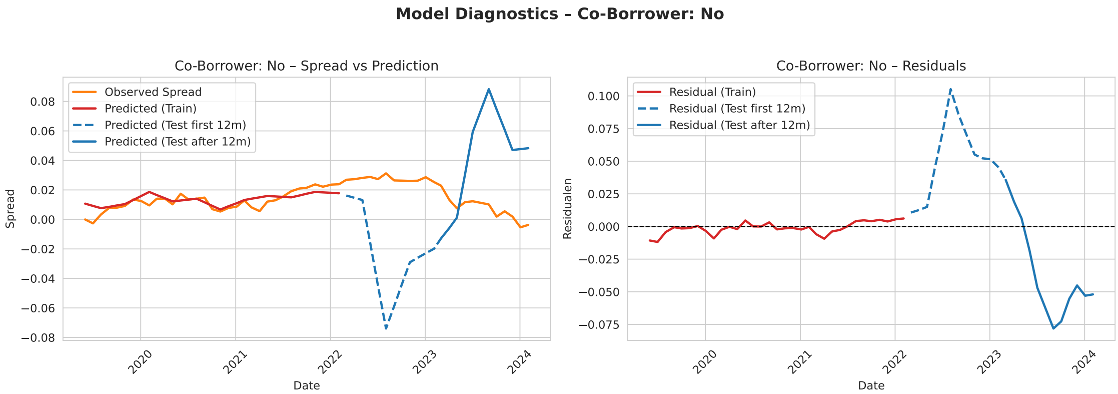

Figure 4.10: The model's predicted spread for the co-borrower segment without a co-borrower, $\mathcal{M}^{(\mathrm{CB,No})}$. The left panel shows the observed spread, $s_t^{(\mathrm{CB,No})}$, and the predicted spread, $\hat{s}_t^{(\mathrm{CB,No})}$, over time. Model performance is poor on the test set. The right panel displays the residuals, $e_t^{(\mathrm{CB,No})}$. Note the different scales on the y-axes.

5

Analysis and Interpretation

5.1 Observed Spread

The observed spread for the total portfolio, $\mathrm{s}_t^{\mathrm{Total}}$, across the sample period, is almost always positive, with a mean of 0.0120 and a maximum of 0.0274 (Table 3.1). This shows that, on average, observed defaults, $\overline{\mathrm{OD}}_t^{\mathrm{Total}}$, have exceeded current PD^{12m}-model predictions, suggesting a structural underestimation of risk in the portfolio-level forecasts. As shown in Figure 3.1, there are noticeable peaks in the spread, especially around the onset of the COVID-19 pandemic. An interpretation of this is that the current $\mathrm{PD}^{\mathrm{12m,\ Norion}}$-model fails to predict defaults accurately during turbulent economic times.

It is also of relevance to acknowledge that the current $\mathrm{PD}^{\mathrm{12m,\ Norion}}$-model is trained on historical data, which is inherently lagged behind real-time developments. This means that the model can never be expected to predict perfectly in times where risk dynamics are changing. Only under relatively stable conditions is it reasonable to expect the model to remain good.

For the age segmentation,$\mathcal{M}^{(\mathrm{Age},\cdot)}$, there is substantial variation in $\mathrm{s}_t^{(\mathrm{Age},\cdot)}$ across borrower age groups (see Figure 3.2). From Table 3.2, we see that the $\mathcal{M}^{(\mathrm{Age,Young})}$ age group exhibits the highest mean spread at 0.0237, compared to 0.0078 for $\mathcal{M}^{(\mathrm{Age,Middle})}$ and 0.0098 for $\mathcal{M}^{(\mathrm{Age,Older})}$. This suggests that the current model underestimation is most severe among younger borrowers. This might be due to greater volatility in their economic conditions and higher sensitivity to macroeconomics. The mid-segmentation has the lowest mean spread, suggesting greater resilience to economic turndowns.

Similarly, spread levels for the mortgage holder segmentation $\mathcal{M}^{(\mathrm{MH},\cdot)}$ vary greatly, see Figure 3.3. Borrowers without a mortgage show the highest mean spread of 0.0280, while those with a mortgage exhibit a negative mean spread -0.0083, indicating potential model overestimation of default risk in the latter group (Table 3.4). In Figure 3.3 we can see that the current $\mathrm{PD}^{\mathrm{12m,\ Norion}}$-model for the Unknown group lies close to zero, suggesting it might be a mixed pool of mortgage and non-mortgage borrowers. This is because mortgage and non-mortgage holder are clearly separated on either side of the horizontal zero level.

There are also notable differences in $\mathrm{s}_t^{(\mathrm{CB},\cdot)}$ depending on co-borrower status (Table 3.4). Loans without a co-borrower exhibit a positive mean spread of 0.0137, indicating that realized defaults have exceeded predictions on average. In contrast, loans with a co-borrower show a negative mean spread of -0.0167, suggesting that the current model has systematically overestimated risk for these borrowers. In Figure 3.4 we can see a huge shift in modeling performance at the beginning of 2023. Norion Bank's PD^{12m}-model is consistently re-evaluated and trained, so this shift could be due to one of those changes in the model.

5.2 Modeling Performance: Comparative Analysis Across Segments

For the total portfolio, $\mathcal{M}^{\text{Total}}$, the AR(1)-corrected model achieves a test MAE of 0.00554 and an R^2 of 0.522, indicating reasonable predictive performance at the portfolio level (Table 4.2). While the training MAE is lower, at 0.00255, the gap between training and test performance indicates some loss of accuracy out-of-sample, though the errors remain low. The plot of predicted and observed spreads in Figure 4.1 shows that the model captures the main fluctuations in the realized spread, though some local deviations remain, especially at the end of the time period. The adjusted prediction, Equation (2.5), outperforms the current model in nine of the twelve months in the final twelve-month unseen test set.

The young portfolio, $\mathcal{M}^{(\text{Age,Young})}$, gap between train and test MAE is 0.00351, meaning the model performed worse on unseen data. We still have a test R^2 of 0.804 showing that the model is generally good. Looking at Figure 4.3 the predicted spread follows the pattern of the observed spread. The adjusted prediction (Equation (2.5)) is better than the current $\overline{\text{PD}}^{\text{12m, Norion}}$ seven out of the final twelve test-months. Forecast errors are small, suggesting that macro signals can explain part of the deviation between observed defaults and current prediction. For the middle portfolio $\mathcal{M}^{(\text{Age,Middle})}$ we can see in Figure 4.4 that the model fails completely on the test set. The reason could be instability in model based on shifts in the chosen features. One of the chosen features are GDPChange with a lag of 30. The model start to fail in September 2022, if you go back 30 months, you end up on Mars 2020, i.e. the start of the COVID-19 pandemic in Sweden. Which had huge downward-impact on the economy, especially GDP. Problems related to the chosen time-frame are discussed further in Chapter 6. Looking at the heatmap correlation plot, Figure B.2, we still have some high values. Further investigation of this showed that no other combination of final features lead to a reasonable model. This model was therefore kept, although having a bit high correlation. For the $\mathcal{M}^{(\text{Age,older})}$ model, raw OLS estimates look reasonable in-sample but fails on out-of-sample data. Only five features were retained from the initial feature selection, indicating that the features might not have the right information to explain the spread in this case. We still have some high correlation between the selected features (Figure B.3). This model was, despite these issues, retained as a trade-off between performance and feature independence. Looking at Figure 4.5 we see that the predicted spread follows the pattern of the observed spread. The model is reasonable but needs some more attention.

Looking at Figure 4.6 for the mortgage-holder group $\mathcal{M}^{(\text{MH,Yes})}$, we see that the model completely fails to generalize on unseen data. Only four features were found in the initial correlation based selection, indicating that, either, there might not be much to gain from the features considered in this thesis, or that the segment itself is not macro-sensitive. Again, GDPChange with lag 30 is involved and might be the reason for the huge shift in prediction. The gap between train and test MAE shows over-fitting (see Table 3.3). The small sample and the narrow, often negative spread, might leave little signal for the macro predictors to detect.

For the $\mathcal{M}^{(\text{MH,No})}$, we see in Figure 4.7, that the predicted spread follows the pattern of the observed spread. Noticeable, R^2 improves from 0.654 to 0.700 on the test set. Ten out of twelve months in the final test period, the adjusted prediction is better than current model. In Figure B.5 we see manageable correlation between features, but this might need more attention. For the $\mathcal{M}^{(\text{MH,Missing})}$, we see in Figure 4.8 that the predicted spread follows the direction of observed spread, but is still not good on unseen data. Despite this, the adjusted prediction beats the current model prediction eleven out of twelve months in final test period. However, the model remains far from reliable.

For the Co-Borrower Yes-segmentation, $\mathcal{M}^{(\text{CB,Yes})}$, the initial correlation-based selection retained only three features. Indicating that there might not be much signal here either. In Figure 4.9 we can see that the model fails on the test data. Looking at Figure 4.10 for the Co-Borrower No-segmentation, $\mathcal{M}^{(\text{CB,No})}$, we also see that it fails on the test set, even though 20 features were initially selected. Neither of these models are fit to use in practice.

5.3 Interpretation of Chosen Macroeconomic Indicators

In Table 4.3 we see the three features that were picked for the total portfolio, all with negative coefficient and in the following order of magnitude: KIX-Index with delta 24; A weaker krona (positive 24-month change) is linked to a narrower spread. Household Confidence, 6-month delta; when confidence improves, the spread narrows. Unemployment Rate, 12-month delta; Rising unemployment reduces the spread. This suggests that the $\text{PD}^{\text{12m, Norion}}$ model already have a strong labor-market effect already, stemming from its target variable. The intercept tells us that even under "average" macro-times, the current model underestimates defaults.

The $\mathcal{M}^{(\text{Age,Young})}$-model's strongest driver is the $\overline{\text{PD}}_t^{\text{12m,Norion(Age,Young)}}$ itself (large negative coefficient), confirming that the underlying prediction already carries a lot of forecasting weight for this sub-group see Table 4.5. A two-year rise in real consumption widens the spread, implying that spending leads to more missed defaults than the $PD^{12m,\text{ Norion}}$ model anticipates — possibly because purchases weaken buffers for the younger households. Again, the large positive intercept shows a systematic baseline under-prediction of defaults even in "normal" times. The $\mathcal{M}^{(\text{Age,Middle})}$-model fails to generalize and is therefore not analyzed further. The only chosen feature for the $\mathcal{M}^{(\text{Age,Older})}$-model is GDPChange with a lag of 30. We can interpret the positive sign as the current $PD^{12m,\text{ Norion}}$ model may assume older borrowers benefit more from good times than what they actually do.

In Table 4.6 we see the final features and their coefficients for the mortgage segmentation. For $\mathcal{M}^{(\text{MH,Yes})}$ the model fails to generalize so no further attention is paid to it. In the $\mathcal{M}^{(\text{MH,No})}$ segmentation, the current PD level again has the greatest magnitude. Additionally, a six-month rise of the household indicator narrows the spread. Stronger GDP momentum over the previous year widens it. Furthermore, we have a large positive intercept showing that even in average conditions, the current PD^{12m}-model understates default risk for borrowers without mortgage. Since the missing segmentation $\mathcal{M}^{(\text{MH,Missing})}$ performs poorly, no additional analysis of coefficients is made.

The Co-Borrower models fail to generalize, so their coefficients (see Table 4.7) should be interpreted with caution. Both models select GDP Change with similar lag lengths of 30 and 26 months respectively. With positive signs suggesting that in stronger economic conditions, the current $PD^{12m,\text{ Norion}}$-model underestimates default risk.

6
Discussion

6.1 Reflection on Methodology

The methodological approach used in this thesis (described in Chapter 3), follows a strict and systematic modeling framework, with the pros of leaving personal bias aside and letting data and algorithms speak for themselves. The cons are that reality might not be optimized for this strict approach - possibly more manual feature selection based on area-specific knowledge, and flexible decisions. The total portfolio, $\mathcal{M}^{\text{Total}}$, performs reasonably well overall but struggles toward the end of the test period, and has some overfitting. The performance in the segmented portfolio varied. We still gain valuable insight into how sensitive different sub-populations were to macroeconomic conditions. Especially, how the spread varies for different segmentation gives clues on how the current $\text{PD}^{\text{12m, Norion}}$-model performs and how it can be improved.

At first, we had to decide whether to treat the problem as as a regression or classification problem. Either modeling portfolio-level defaults, or individual contract defaults as the target. The latter was investigated using XGBoost and logistic regression classifiers, neither performed satisfactorily. Additionally, Norion Bank would, for internal modeling policy reasons, not implement a classification solution. Hence, the regression approach was chosen.

Linear Regression (LR) was picked based on the easy interpretability of the final model and its coefficients. In hindsight, LR, is likely too rigid to capture the dynamics of macroeconomic shifts and its impact on credit risk. The model is also sensitive to how the train/test-split is made. If the train period only see turbulent macroeconomic conditions, then it will most likely not perform well on a test period that is calmer, and vice versa. A result of picking LR was also that we got limited data points. Due to fundamental changes in Norion Bank's $\text{PD}^{\text{12m, Norion}}$ model it only made sense to consider data after 2019-06. Also, due to the 12-month realization delay of default, the end date had to be set at 2024-02. Making a set of only 56 data points. A combination of this and structural breaks and volatility (e.g., GDP crash in test data) may have distorted parts of the model, resulting in poor test performance. Further, long training windows may dilute temporal sensitivity - rolling windows might be an option. This approach was briefly explored with no immediate satisfactory result. The current $\text{PD}^{\text{12m, Norion}}$ model of Norion Bank do not explicitly incorporate macroeconomic features, but this does not mean that such information is not partly captured through the target variable and borrower-specific information.

6.2 Segment-wise Model Insights and Takeaways

The total portfolio model, $\mathcal{M}^{\text{Total}}$, performed well overall, choosing a small set of macroeconomic features that explains the spread well. Showing that the persistent underestimation of default risk can be modeled by shifts in exchange rate, consumer confidence, and unemployment. However, some overfitting is present which needs to be handled.

Model performance varied notably across the borrower segmentation. The models for young borrowers $\mathcal{M}^{(\text{Age,Young})}$, older borrowers $\mathcal{M}^{(\text{Age,Older})}$, and no-mortgage group $\mathcal{M}^{(\text{MH,No})}$, generalized well, with low test errors and interpretable macro effects. Most models still have a problem with overfitting. In contrast, the model for middle-age borrowers $\mathcal{M}^{(\text{Age,Middle})}$, borrowers with a mortgage $\mathcal{M}^{(\text{MH,Yes})}$, and single-borrowers $\mathcal{M}^{(\text{CB,No})}$, suffered from bad generalization and overfitting.

Macro indicators had different relevance across groups. For young borrowers, $\mathcal{M}^{(\text{Age,Young})}$, consumption growth influenced prediction default gaps, while older borrowers,$\mathcal{M}^{(\text{Age,Older})}$, were more sensitive to long-lagged GDP trends. In the no-mortgage group,$\mathcal{M}^{(\text{MH,No})}$, household confidence and GDP growth explained the model error. Segments with poor performance either lacked stable macro relationships or were disproportionately affected by large macro shifts (e.g., GDP crash due to COVID-19), showing the limits of linear models.

6.3 Practical Implications for Norion Bank and IFRS 9 Modeling

While some results show that macroeconomic variables can help with explaining the deviations in the current model, it is likely too early to apply these effects as exogenous adjustments. There is not enough stability in the models predictions, and overfitting is still an issue. However, the framework and findings can definitely be learned from and used as a starting point for further exploration. By identifying when and where PD models tend to under- or overestimate default risk, Norion can also use it to enhance awareness of the current $\text{PD}^{\text{12m, Norion}}$-model limitations, informing the design of current and future models, or motivate recalibration of specific segments.

6.4 Limitations of Study and Results

There are several limitations that affect the interpretation and generalization of the results. Firstly, the current $\text{PD}^{\text{12m, Norion}}$-model already seems to indirectly, through borrower-level information, capture some macroeconomic effects. This makes it difficult to isolate the added value of macro variables without risking double-counting. Secondly, the limited number of time points means that the model is trained over a narrow range of macroeconomic conditions. This restricts the model's ability to generalize to unseen macroeconomic regimes. As mentioned before, the extreme changes around COVID-19 did affect the performance. In particular, a few extreme observations (such as the GDP drop during COVID-19) can disproportionately influence results, leading to instability or overfitting. Third, there are some data-related issues including, interpolation, revision of data, and date-mismatch, see subsection 3.2. There are also seasonal effects that might influence the results. Previous work at Norion Bank points to a seasonal affect on defaults across the year. Then, having a training window that is not in full-years, could affect the model.

6.5 Suggestions for Future Research

Several directions could build on this work to address current limitations and strengthen both the explanatory and model stability issues:

- Nonlinear modeling: as discussed above, there are limitations with the linear regression-modeling approach. The relationships between macro indicators and the spread might be more complex and require non-linear models like decision trees.
- Exploring other adjustment terms: with this thesis post-hoc exogenous approach, the macroeconomic correction is added linearly to the current $\text{PD}^{\text{12m, Norion}}$ model (see Equation (**??**)). But, there is no reason the correction has to be an additive term. One could instead try a multiplicative factor, or even define the correction as a function involving multiple terms or non-linear transformations. This might capture more complex relationships between macroeconomic variables and the spread.
- Alternative target: instead of modeling the spread, one could try to model the defaults directly, i.e. use $\overline{OD}_t^{12m}$ as a target. Maybe, macroeconomic conditions solely can accurately predict the default rates.
- Capping test data: as discussed before, there are big issues in linear modeling when the training and test data differs. One approach here is to cap the test data values, based on the distribution of the training data. Not allowing extreme values to influence the predictions. One could make a "hard-cap" not allowing test-values in macroeconomic variables to exceed the training values, or have some linear interpolation between the two values in test/train data. This was briefly explored and the conclusion is that there were some promising results. Models with capped test data showed more stable performance and were less prone to complete failure on the test set.
- Macro integration into the PD model itself: whether incorporating the macroeconomic variables as a feature into the current classification model and see if it improves predictions. Though, as mentioned above, this was not done as there are currently reasons limiting the possibility of implementing such a model at Norion Bank due to how the other model are designed.

Bibliography

[1] Gustaver, M. (2020) A Chalmers University of Technology Master's thesis template for LaTeX. Unpublished.

[2] IFRS Foundation (2014). *IFRS 9 Financial Instruments*. International Accounting Standards Board.

[3] Pesaran, M. H., Schuermann, T., Treutler, B. J., Weiner, S. M. (2006). Macroeconomic dynamics and credit risk: A global perspective. Journal of Money, Credit and Banking, 38(5), 1211–1261.

[4] Abad, J., Suarez, J. (2018). The procyclicality of risk-based capital requirements: Evidence from the euro area. European Systemic Risk Board (ESRB) Occasional Paper Series, No. 12.

[5] European Banking Authority (2017). IFRS 9 and the EBA Guidelines on credit risk and accounting for expected credit losses. EBA Report.

[6] Altman, E. I., Saunders, A. (2002). Credit risk measurement: Developments over the last 20 years. Journal of Banking Finance, 21(11–12), 1721–1742.

[7] Statistics Sweden (SCB). (2024). *Consumer Price Index (CPI) – Inflation in Sweden*. Available at: `https://www.scb.se/en/finding-statistics/statistics-by-subject-area/prices-and-consumption/consumer-price-index/`

[8] Sveriges Riksbank. (2024). *Monetary policy and policy rate*. Available at: `https://www.riksbank.se/en-gb/monetary-policy/the-inflation-target-and-the-interest-rate/`

[9] Louzis, D. P., Vouldis, A. T., Metaxas, V. L. (2012). Macroeconomic and bank-specific determinants of non-performing loans in Greece: A comparative study of mortgage, business and consumer loan portfolios. Journal of Banking Finance, 36(4), 1012–1027.

[10] Koopman, S. J., Kräussl, R., Lucas, A., Monteiro, A. (2009). Credit cycles and macro fundamentals. Journal of Empirical Finance, 16(1), 42–54.

[11] Sveriges Riksbank. (2024). *KIX Index – Definition and Weights*. Available at: `https://www.riksbank.se/en-gb/statistics/search-interest--exchange-rates/exchange-rates/kix-index/`

[12] National Institute of Economic Research (NIER). (2024). *Economic Sentiment Indicators*. Available at: `https://www.konj.se/english/statistics/economic-tendency-survey.html`

[13] Kuhn, M., Johnson, K. (2019). *Feature Engineering and Selection: A Practical Approach for Predictive Models*. CRC Press.

[14] Hastie, T., Tibshirani, R., Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.

[15] Wasserman, L. (2013). *All of Statistics: A Concise Course in Statistical Inference*. Springer.

[16] Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5th ed.). Cengage Learning.

[17] Greene, W. H. (2018). *Econometric Analysis* (8th ed.). Pearson Education.

[18] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: With Applications in R*. Springer.

[19] Greene, W. H. (2012). *Econometric Analysis (7th ed.)*. Pearson Education.

[20] Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4), 591–611.

[21] Godfrey, L. G. (1978). Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica*, 46(6), 1293–1301.

[22] Durbin, J., & Watson, G. S. (1951). Testing for serial correlation in least squares regression. *Biometrika*, 38(1–2), 159–178.

[23] Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47(5), 1287–1294.

[24] Cochrane, D., & Orcutt, G. H. (1949). Application of least squares regression to relationships containing autocorrelated error terms. *Journal of the American Statistical Association*, 44(245), 32–61.

A

Appendix 1

In this appendix we summarize the macroeconomic indicators used in the study in Table A.1 (Next side).

Table A.1: Macroeconomic Indicators Used in the Study

Category	Indicator	Description	Source	Unit	Freq.
Inflation	KPIF Fixed Baseline	Consumer Price Index excluding mortgage rate effects	SCB	Index	Monthly
	KPIF	Inflation rate measured as year-over-year change	SCB	% YoY	Monthly
Labor Market	Unemployment Rate	Percentage of labor force unemployed, seasonally adjusted	SCB	%	Monthly
Monetary Policy	Repo Rate	Central bank's policy interest rate	Riksbank	%	5-6 times/year
Economic Growth	GDP Growth	Seasonally adjusted percentage change in GDP	SCB	% QoQ	Quarterly
Consumption	Household Consumption	Total household expenditure in the economy	SCB	SEK (millions)	Monthly
Economic Sentiment	Economic Barometer	Overall sentiment index across economic sectors	NIER	Index	Monthly
	Business Barometer	Confidence level among businesses	NIER	Index	Monthly
	Retail Trade Indicator	Confidence in the retail sector	NIER	Index	Monthly
	Service Sector Indicator	Confidence in the service industry	NIER	Index	Monthly
	Consumer Confidence	Household perceptions of their financial situation	NIER	Index	Monthly
	Household Macro Index	Summary measure of macroeconomic expectations	NIER	Index	Monthly
Exchange Rate	KIX Index	Trade-weighted exchange rate of the Swedish krona	Riksbank	Index	Daily

B

Appendix 2

In this Appendix, we gather additional results for the different segmentation as Tables and Figures.

B.1 Additional Result on Age Segmentation, $\mathcal{M}^{(\mathbf{Age},\cdot)}$

B.1.1 Young, $\mathcal{M}^{(\mathbf{Age},\mathbf{Young})}$

This subsection presents the additional results for the Young age segment. It includes the selected features from ElasticNet, correlation heatmap, regression model performance (OLS and AR-corrected), and residual diagnostics.

Table B.1: Features selected by ElasticNet for the Young Age Group, $\mathcal{M}^{(\text{Age},\text{Young})}$. Features are ordered by magnitude of their coefficient.

Feature	**Coefficient**	**Sign**
MeanPredictionPd12m	-6.54×10^{-3}	–
ServiceSectorIndic_lag_6	2.85×10^{-3}	+
ServiceSectorIndic_roll_12	1.66×10^{-3}	+
ConfIndicHousehold_lag_6	8.66×10^{-4}	+
Consumption_delta_24	7.63×10^{-4}	+
KpifFixedBaseLine_delta_24	2.66×10^{-4}	+
Consumption_roll_6	2.20×10^{-5}	+

Table B.2: Model performance – OLS and AR(1)-corrected Young Age Group, $\mathcal{M}^{(\text{Age},\text{Young})}$. OLS-model is picked as final based on lower train MAE.

Metric	**Train**		**Test**	
	OLS	AR(1)	OLS	AR(1)
MAE	0.00332	0.00419	0.00683	0.00872
R^2	0.890	0.855	0.804	0.782
Adjusted R^2	0.883	0.845	0.784	0.760
Durbin–Watson	0.508	0.295	—	—

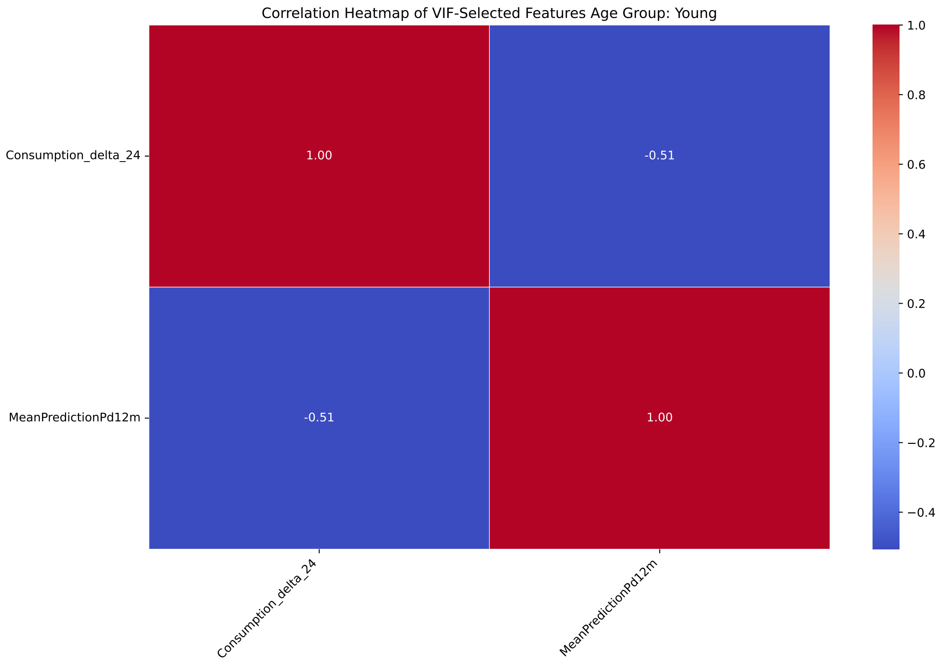

Figure B.1: Correlation heatmap for the final features in $\mathcal{M}^{(\text{Age,Young})}$.

Table B.3: Summary of residual diagnostic tests (retained OLS model) on $\mathcal{M}^{(\text{Age,Young})}$.

Test	**Statistic**	**p-value**
Shapiro–Wilk (normality)	0.9723	0.5459
Breusch–Pagan (LM test)	13.3948	0.0012
Breusch–Pagan (F test)	10.2484	0.0004
Breusch–Godfrey (LM test)	19.1815	0.0001
Breusch–Godfrey (F test)	19.4334	< 0.0001

B.1.2 Middle, $\mathcal{M}^{(\text{Age,Middle})}$

Below are the extended results for the Middle-aged borrower segment. Tables and figures detail selected features, correlation heatmaps, model performance under both OLS and AR(1) correction, as well as results from residual diagnostic testing.

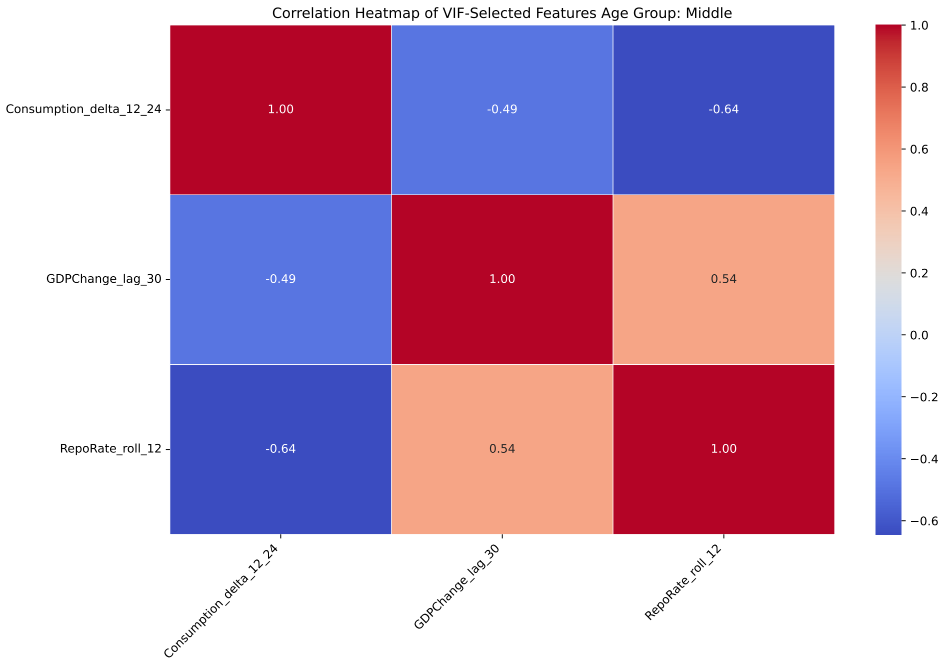

Figure B.2: Correlation heatmap for the final features in $\mathcal{M}^{(\text{Age,Middle})}$.

Table B.4: Features selected by ElasticNet Middle-Age Segment, $\mathcal{M}^{(\text{Age,Middle})}$. Features are ordered on the magnitude of their coefficients.

Feature	**ElasticNet Coefficient**	**Sign**
MeanPredictionPd12m	-1.01×10^{-2}	–
RepoRate_roll_12	-4.63×10^{-3}	–
Consumption_roll_12	3.23×10^{-3}	+
EconomicBarometer_lag_25	2.35×10^{-3}	+
Kpif_lag_23	1.22×10^{-3}	+
KixIndex_lag_22	8.80×10^{-4}	+
Consumption_lag_26	8.56×10^{-4}	+
HousholdMakroIndex_lag_20	-5.81×10^{-4}	–
GDPChange_lag_30	-5.17×10^{-4}	–
KpifFixedBaseLine_lag_6	5.41×10^{-4}	+
RepoRate_lag_30	-2.98×10^{-4}	–
ConfIndicHousehold_lag_17	2.07×10^{-4}	+
UnemploymentProcent_lag_16	-1.63×10^{-4}	–
Consumption_delta_12_24	-6.10×10^{-5}	–

Table B.5: Model performance – OLS and AR(1)-corrected Middle-Age Segment, $\mathcal{M}^{(\text{Age,Middle})}$. OLS-model picked as final based on lower train MAE.

Metric	**Train**		**Test**	
	OLS	AR(1)	OLS	AR(1)
MAE	0.00375	0.00412	0.04974	0.03727
R^2	0.649	0.585	-38.331	-20.185
Adjusted R^2	0.626	0.558	-42.264	-22.304
Durbin–Watson	0.919	0.687	—	—

Table B.6: Summary of diagnostic tests. OLS-Model, Middle-Age Segment, $\mathcal{M}^{(\text{Age,Middle})}$.

Test	**Statistic**	**p-value**
Shapiro–Wilk (normality)	0.9681	0.4293
Breusch–Pagan (LM test)	4.0885	0.1295
Breusch–Pagan (F test)	2.1212	0.1375
Breusch–Godfrey (LM test)	11.2244	0.0037
Breusch–Godfrey (F test)	7.2165	0.0030

B.1.3 Older, $\mathcal{M}^{(\text{Age,Older})}$

This subsection summarizes the results for the Older age segment. It covers ElasticNet-selected features, correlation analysis, comparative model performance (OLS vs. AR-corrected), and residual diagnostics.

Table B.7: Features selected by ElasticNet Older Age Segment, $\mathcal{M}^{(\text{Age,Older})}$. Features ordered by magnitude of their coefficients.

Feature	**ElasticNet Coefficient**	**Sign**
MeanPredictionPd12m	-3.13×10^{-3}	−
Consumption_lag_10	2.25×10^{-3}	+
RepoRate_roll_3	3.13×10^{-4}	+
GDPChange_lag_30	5.15×10^{-4}	+
RepoRate_lag_2	-1.47×10^{-4}	−

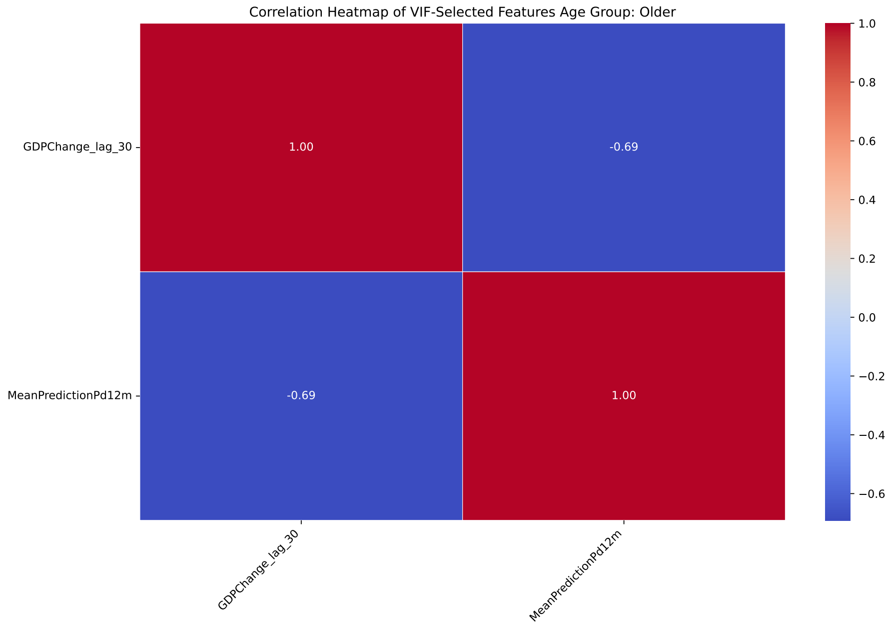

Figure B.3: Correlation heatmap of selected features in the Older Age Segment, $\mathcal{M}^{(\text{Age,Older})}$. Darker shades indicate stronger (positive or negative) correlations.

Table B.8: Model performance – OLS and AR(1)-corrected, Older Age Segment, $\mathcal{M}^{(\text{Age,Older})}$. OLS-model picked as final based on lower train MAE.

Metric	**Train**		**Test**	
	OLS	AR(1)	OLS	AR(1)
MAE	0.00295	0.00475	0.02172	0.00445
R^2	0.631	0.208	-16.588	0.424
Adjusted R^2	0.607	0.156	-18.347	0.366
Durbin–Watson	0.488	0.070	—	—

Table B.9: Summary of diagnostic tests. OLS-Model, Older Age Segment,$\mathcal{M}^{(\text{Age,Older})}$.

Test	**Statistic**	**p-value**
Shapiro–Wilk (normality)	0.9606	0.2696
Breusch–Pagan (LM test)	1.3870	0.4998
Breusch–Pagan (F test)	0.6581	0.5251
Breusch–Godfrey (LM test)	18.9651	0.0001
Breusch–Godfrey (F test)	18.9179	< 0.0001

B.2 Additional results on Mortgage Holder segmentation, $\mathcal{M}^{(\mathbf{MH},\cdot)}$

B.2.1 Yes, $\mathcal{M}^{(\mathbf{MH},\mathbf{Yes})}$

This subsection presents additional results for the segment of borrowers with a mortgage. It includes selected features, a correlation heatmap, model performance for both OLS and AR(1)-corrected models, and diagnostic test for the residuals.

Table B.10: Features selected by ElasticNet, Mortgage Segment: Yes, $\mathcal{M}^{(\mathrm{MH},\mathrm{Yes})}$. Features ordered by magnitude of their coefficients.

Feature	**ElasticNet Coefficient**	**Sign**
KpifFixedBaseLine_lag_11	3.58×10^{-4}	+
Consumption_lag_26	2.42×10^{-3}	+
RepoRate_roll_3	1.75×10^{-3}	+
GDPChange_lag_30	2.26×10^{-3}	+
MeanPredictionPd12m	-4.88×10^{-3}	−

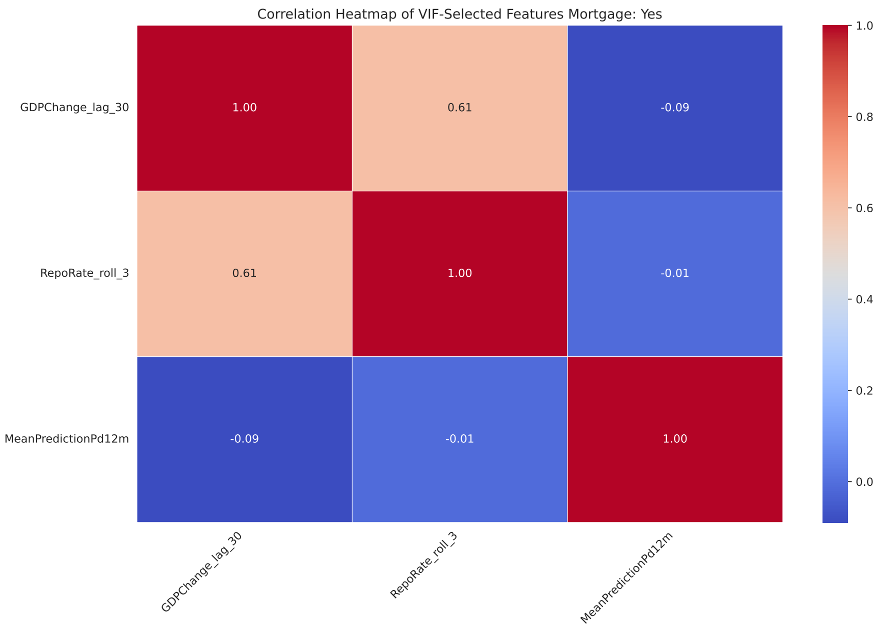

Figure B.4: Correlation heatmap of selected features in the Mortgage Segment: Yes, $\mathcal{M}^{(\mathrm{MH},\mathrm{Yes})}$. Stronger colors represent higher absolute correlation values.

Table B.11: Model performance – OLS and AR(1)-corrected, Mortgage Segment: Yes, $\mathcal{M}^{(\text{MH,Yes})}$. OLS-model picked as final based on better train MAE.

Metric	Train		Test	
	OLS	AR(1)	OLS	AR(1)
MAE	0.00305	0.00332	0.05886	0.04216
R^2	0.811	0.651	–118.532	–50.603
Adjusted R^2	0.792	0.615	–137.405	–58.751
Durbin–Watson	0.964	0.470	—	—

Table B.12: Summary of diagnostic tests. OLS-model, Mortgage Segment: Yes, $\mathcal{M}^{(\text{MH,Yes})}$.

Test	Statistic	p-value
Shapiro–Wilk (normality)	0.822	< 0.001
Durbin–Watson (autocorrelation)	0.470	—
Breusch–Pagan (LM test)	13.80	0.0032
Breusch–Pagan (F test)	6.94	0.0012
Breusch–Godfrey (LM test)	8.70	0.0129
Breusch–Godfrey (F test)	4.83	0.0161

B.2.2 No, $\mathcal{M}^{(\text{MH,No})}$

This subsection shows detailed results for borrowers without a mortgage. Tables report the selected macro features, regression performance metrics, and the results of residual diagnostic testing.

Table B.13: Features selected by ElasticNet, Mortgage: No, $\mathcal{M}^{(\text{MH,No})}$. Features ordered by the magnitude of their coefficients.

Feature	ElasticNet Coefficient	Sign
MeanPredictionPd12m	-2.22×10^{-3}	–
ConfIndicHousehold_delta_6	-1.26×10^{-3}	–
Consumption_lag_6	1.27×10^{-3}	+
GDPChange_lag_5	1.23×10^{-3}	+
ServiceSectorIndic_lag_6	1.19×10^{-3}	+
GDPChange_roll_12	1.98×10^{-4}	+

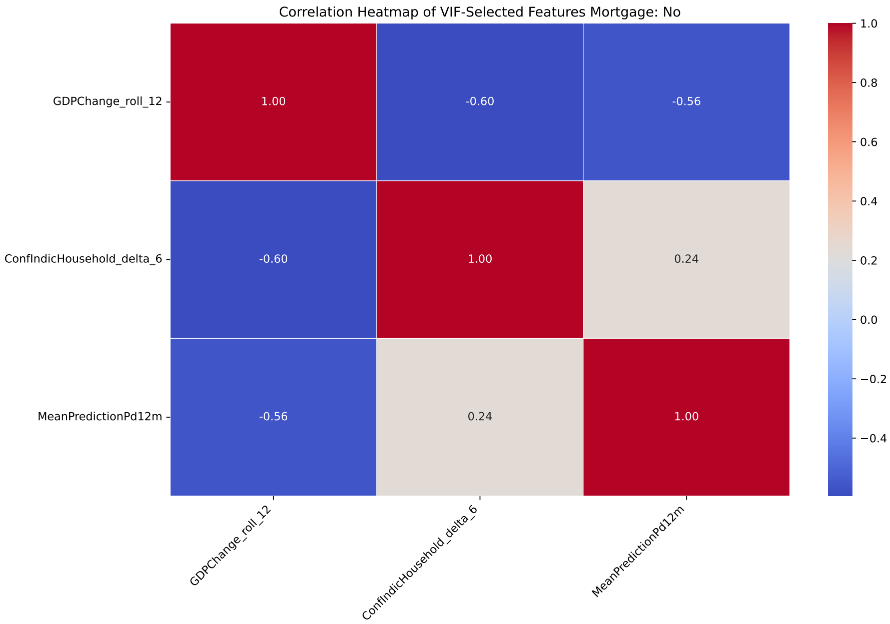

Figure B.5: Correlation heatmap of selected features in the Mortgage Segment: No, $\mathcal{M}^{(\mathrm{MH,No})}$. Stronger colors indicate higher absolute correlations.

Table B.14: Model performance – OLS and AR(1)-corrected, Mortgage: No, $\mathcal{M}^{(\mathrm{MH,No})}$. OLS-model picked as final based on lower train MAE.

Metric	**Train**		**Test**	
	OLS	AR(1)	OLS	AR(1)
MAE	0.00348	0.00431	0.00635	0.00710
R^2	0.654	0.226	0.700	0.719
Adjusted R^2	0.619	0.146	0.652	0.675
Durbin–Watson	0.977	0.280	—	—

Table B.15: Summary of diagnostic tests, Mortgage: No, $\mathcal{M}^{(\mathrm{MH,No})}$, OLS-model.

Test	**Statistic**	**p-value**
Shapiro–Wilk (normality)	0.9578	0.2245
Breusch–Pagan (LM test)	1.7144	0.6337
Breusch–Pagan (F test)	0.5297	0.6655
Breusch–Godfrey (LM test)	7.3950	0.0248
Breusch–Godfrey (F test)	3.8990	0.0325

B.2.3 Unknown, $\mathcal{M}^{(\mathrm{MH,Unknown})}$

This subsection contains additional results for contracts where mortgage status is unknown. It includes the ElasticNet feature set, correlation heatmap, model evaluation for OLS and AR(1), and residual diagnostics.

Table B.16: Features selected by ElasticNet, Mortgage: Unknown, $\mathcal{M}^{(\mathrm{MH,Unknown})}$. Features ordered by the magnitude of their coefficients.

Feature	ElasticNet Coefficient	Sign
MeanPredictionPd12m	-7.98×10^{-3}	−
KpifFixedBaseLine_roll_12	-5.07×10^{-3}	−
EconomicBarometer_lag_26	3.62×10^{-3}	+
Consumption_roll_12	3.23×10^{-3}	+
ConfIndicHousehold_lag_6	1.68×10^{-3}	+
Consumption_lag_4	1.38×10^{-3}	+
UnemploymentProcent_lag_18	1.01×10^{-3}	+
KixIndex_delta_12_24	4.79×10^{-4}	+
GDPChange_lag_26	2.80×10^{-5}	+

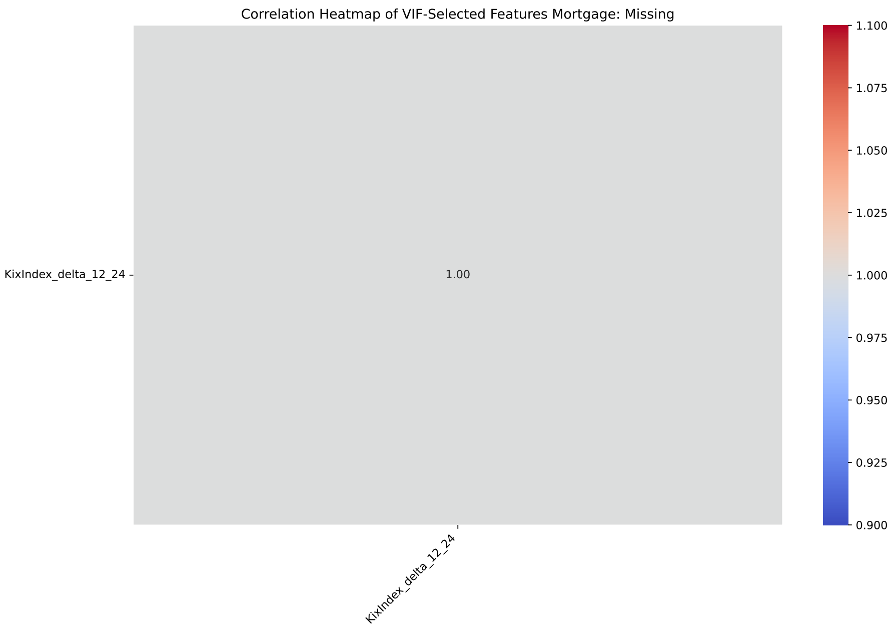

Figure B.6: Correlation heatmap of selected features in the Mortgage Segment: Unknown, $\mathcal{M}^{(\mathrm{MH,Unknown})}$. Stronger colors indicate higher absolute correlations.

Table B.17: Model performance – OLS and AR(1)-corrected, Mortgage: Unknown, $\mathcal{M}^{(\mathrm{MH,Unknown})}$. AR(1)-corrected model picked as final based on better train MAE.

Metric	Train		Test	
	OLS	AR(1)	OLS	AR(1)
MAE	0.00264	0.00263	0.00827	0.00792
R^2	0.516	0.508	-1.470	-1.267
Adjusted R^2	0.501	0.492	-1.587	-1.375
Durbin–Watson	0.960	0.940	—	—

Table B.18: Summary of diagnostic tests, Mortgage: Unknown, $\mathcal{M}^{(\mathrm{MH,Unknown})}$, AR(1)-corrected model.

Test	Statistic	p-value
Shapiro–Wilk (normality)	0.9775	0.7083
Breusch–Pagan (LM test)	2.9019	0.0885
Breusch–Pagan (F test)	2.9889	0.0938
Breusch–Godfrey (LM test)	7.0176	0.0299
Breusch–Godfrey (F test)	3.9163	0.0312

B.3 Additional results on Co-Borrower status, $\mathcal{M}^{(\mathbf{CB},\cdot)}$

B.3.1 Yes, $\mathcal{M}^{(\mathbf{CB,Yes})}$

This subsection reports the results for borrowers with a co-borrower. The tables include selected macro features be ElasticNet, correlation heatmap, regression performance (OLS and AR-corrected), and residual diagnostics.

Table B.19: Features selected by ElasticNet, Co-Borrower: Yes, $\mathcal{M}^{(\mathrm{CB,Yes})}$. Features ordered by magnitude of their coefficients.

Feature	ElasticNet Coefficient	Sign
MeanPredictionPd12m	-4.44×10^{-3}	–
Consumption_lag_3	1.14×10^{-3}	+
GDPChange_lag_30	3.86×10^{-4}	+

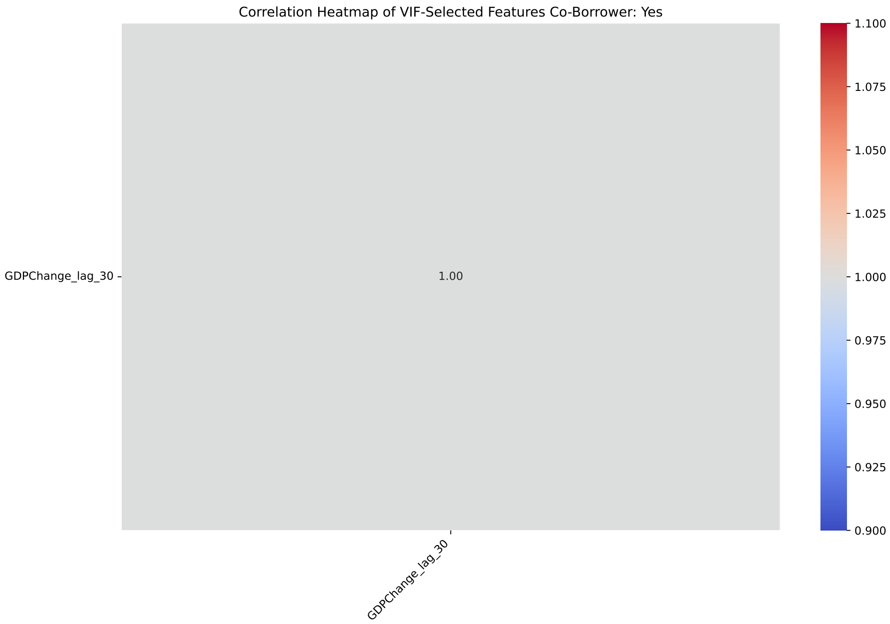

Figure B.7: Correlation heatmap of selected features in the Co-Borrower: Yes segment, $\mathcal{M}^{(\mathrm{CB,Yes})}$. Stronger colors indicate higher absolute correlations.

Table B.20: Model performance – OLS and AR(1)-corrected. Co-Borrower: Yes, $\mathcal{M}^{(\mathrm{CB,Yes})}$. AR(1)-corrected model picked as final based on lower train MAE.

Metric	**Train**		**Test**	
	OLS	AR(1)	OLS	AR(1)
MAE	0.00366	0.00355	0.04612	0.03598
R^2	0.523	0.483	-84.299	-49.450
Adjusted R^2	0.508	0.467	-88.361	-51.853
Durbin–Watson	1.192	1.105	—	—

Table B.21: Summary of diagnostic tests, Co-Borrower: Yes,$\mathcal{M}^{(\mathrm{CB,Yes})}$, OLS model.

Test	**Statistic**	**p-value**
Shapiro–Wilk (normality)	0.9236	0.0230
Breusch–Pagan (LM test)	6.6046	0.0102
Breusch–Pagan (F test)	7.7567	0.0090
Breusch–Godfrey (LM test)	4.4723	0.1069
Breusch–Godfrey (F test)	2.2732	0.1210

B.3.2 No, $\mathcal{M}^{(\mathbf{CB},\mathbf{No})}$

This subsection provides additional results for borrowers without a co-borrower. It includes the ElasticNet-selected features, correlation heatmap, regression model comparison, and diagnostic statistics.

Table B.22: Features selected by ElasticNet Co-Borrower: No, $\mathcal{M}^{(\mathrm{CB},\mathrm{No})}$. Features ordered by magnitude of their coefficients.

Feature	**ElasticNet Coefficient**	**Sign**
MeanPredictionPd12m	-6.12×10^{-3}	$-$
Consumption_roll_12	2.30×10^{-3}	$+$
EconomicBarometer_lag_26	2.15×10^{-3}	$+$
ConfIndicHousehold_lag_6	2.02×10^{-3}	$+$
Consumption_delta_12_24	7.95×10^{-4}	$+$
Kpif_delta_24	-7.17×10^{-4}	$-$
GDPChange_lag_26	4.07×10^{-4}	$+$
Consumption_lag_4	4.96×10^{-4}	$+$
BarometerBusinessTotal_lag_25	2.82×10^{-4}	$+$
ConfIndicHousehold_lag_18	4.00×10^{-5}	$+$

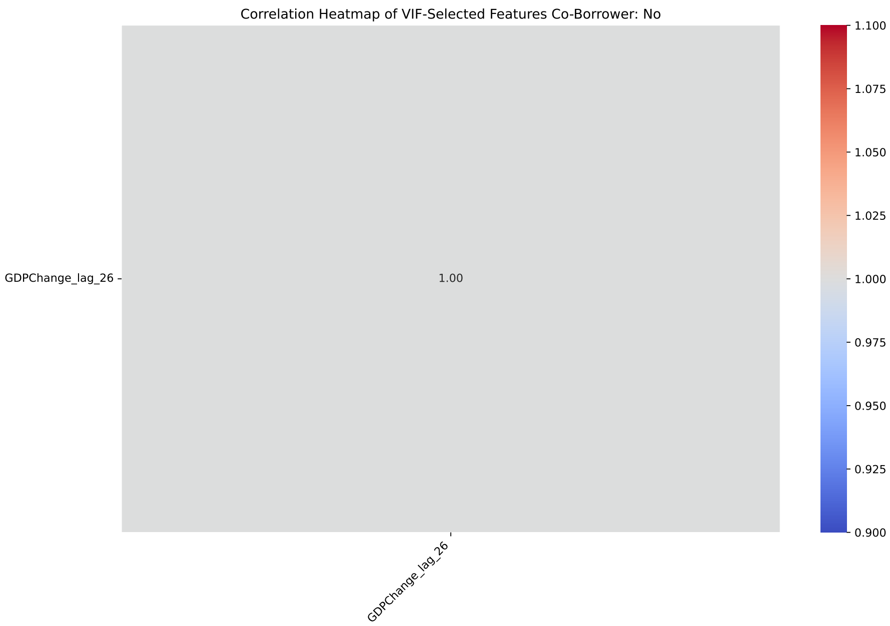

Figure B.8: Correlation heatmap of selected features in the Co-Borrower: No segment, $\mathcal{M}^{(\mathrm{CB},\mathrm{No})}$. Stronger colors indicate higher absolute correlations.

Table B.23: Model performance – OLS and AR(1)-corrected, Co-Borrower: No, $\mathcal{M}^{(\mathrm{CB,No})}$. OLS-model is picked as final based on lower train MAE.

Metric	**Train**		**Test**	
	OLS	AR(1)	OLS	AR(1)
MAE	0.00337	0.00359	0.06578	0.04796
R^2	0.570	0.487	-41.242	-21.097
Adjusted R^2	0.557	0.471	-43.254	-22.149
Durbin–Watson	0.729	0.532	—	—

Table B.24: Summary of diagnostic tests, Co-Borrower: No, OLS-model)

Test	**Statistic**	**p-value**
Shapiro–Wilk (normality)	0.9015	0.0058
Breusch–Pagan (LM test)	1.0006	0.3172
Breusch–Pagan (F test)	0.9693	0.3325
Breusch–Godfrey (LM test)	11.8659	0.0027
Breusch–Godfrey (F test)	8.1412	0.0016

DEPARTMENT OF SOME SUBJECT OR TECHNOLOGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY