MSc Finance – Grade 8.5/10

The application of the LASSO machine learning technique as model selection method of macroeconomic variables for the Point-in-Time Probability of Default adjustment within the IFRS 9 framework

Author: Thomas van den Berg
Date: June 17, 2024
Supervisor: Dr. X. Wang

Abstract

The International Financial Reporting Standard 9 (IFRS 9) regulatory framework obliges banks to calculate Expected Credit Losses (ECL) on lending portfolios, requiring a transition from Through-the-Cycle (TtC) Probability of Default (PD) to Point-in-Time (PiT) PD. For this PiT PD adjustment, it is essential to identify the most explanatory macroeconomic variables (MEVs) for the default rate time series. However, no prescribed method for the selection currently exists. To address this gap, I propose the LASSO BIC method, an automated machine learning technique for efficient selection of the MEVs. Using European corporate default rate data from Moody's, I apply the LASSO BIC method and demonstrate its effectiveness in selecting significant MEVs, with the 10th lag of the first differenced 3M Euribor rate and the 8th lag of the first differenced European corporate bond spread consistently identified as the most important variables for all forecasted years, except for 2019. The LASSO BIC method is a more efficient approach compared to the existing literature. The results indicate that the explanatory power of selected MEVs vary over time, indicating the need for periodic re-evaluation of the selected MEVs. Although frequent updates of MEVs in a bank's PiT PD adjustments are impractical, banks should consider frequent re-examination of their selected MEVs as new default rate data becomes available.

Keywords: IFRS 9 | European corporate loan default rate | Machine learning | OLS regression | Forecasting | LASSO BIC | Point-in-Time Probability of Default | Expected Credit Losses

A thesis submitted in fulfilment of the requirements for the VU Master of Science degree in Finance. In this research, I used Generative Artificial Intelligence (GenAI) to efficiently search for the relevant literature and, to a limited extent, for more efficient coding purposes.

Amsshare

Table of Contents

Chapter 1: Introduction *1*

1.1 Reading guide 3

Chapter 2: Literature review *4*

2.1 Regulatory details 4

2.2 Existing literature 6

2.3 Complexity theories in economics 10

Chapter 3: Methodology and Data *11*

3.1 Macro-economic determinants of default rates 11

3.1.1 Euribor 3M interest rate 11

3.1.2 European Industrial Production (GDP) 11

3.1.3 European Harmonised Index Consumer Prices (HICP) 12

3.1.4 European Traded Volume (XVOL) 12

3.1.5 European Unemployment rate 13

3.1.6 European Governmental Bond Spread 13

3.1.7 European Corporate Bond Spread 13

3.2 IFRS 9 Point-in-Time methodology 14

3.3 LASSO model selection methodology 20

3.3.1 Data transformations 22

3.4 Forecasting methodology 22

3.5 Forecast evaluation 25

Chapter 4: Results *26*

4.1 The LASSO BIC selected models 27

4.2 Conclusions on hypotheses 30

4.3 Forecast results 32

Chapter 5: Conclusion *35*

5.1 Limitations 35

References *37*

Appendices *43*

Chapter 1: Introduction

The IFRS 9 framework, an international regulatory standard, requires banks to calculate the Expected Credit Losses (ECL) on their lending portfolios. According to the IFRS 9 regulation, the Probability of Default (PD) should be adjusted from a Through-the-Cycle (TtC) PD to a Point-in-Time (PiT) PD. Therefore, macroeconomic variables, further named MEVs, need to be selected and their forecasts need to be incorporated to compute the PiT PD. However, there is no clear prescribed method for banks on how to select these MEVs (Basel Committee on Banking Supervision, 2015). According to the report of the European Banking Authority (EBA), most European banks use GDP, while others also consider additional MEVs such as the unemployment rate, inflation and the 3M Euribor (the 3-month European interbank interest rate). Regression analysis is a common method to select the macroeconomic variables for the PiT PD adjustment (Durović, 2019; Bocchio et al., 2023). Breed et al. (2023) propose the Principal Component Regression as methodology to select the MEVs. While these methods are commonly used in the industry, their primary drawback is the need to perform numerous combinations of the MEVs and their lags. This process is highly labour-intensive and presents a major disadvantage. To address this issue, I propose the automated model selection method named LASSO BIC. The LASSO BIC machine learning technique offers automated variable selection, meaning computational efficiency and better performance on high dimensional data (Angrist & Frandsen, 2022). Thus, can the LASSO BIC method select the most explanatory MEVs as a more efficient method?

To answer this question, I perform the LASSO BIC machine learning model selection method on European aggregated corporate loan default rate data extracted from Moody's database (2024) and test the forecast performance of these models for five forecasted years. I find that the LASSO BIC method efficiently selects the significant MEVs. Furthermore, the LASSO BIC method consistently selects the 10^{th} lag of the first differenced 3M Euribor rate and the 8^{th} lag of the first differenced European corporate bond spread as significant explanatory variables for all forecasted years, except for the 2019 forecast, using an expanding estimation window. As the estimation windows vary, so do the selected macroeconomic variables, reflecting changes in their explanatory power over time.

Before applying the LASSO BIC selection method, the number of lags for the macroeconomic variables included in the analysis must be determined. Furthermore, both the default rate time series (dependent variable) as well as the MEVs (independent variables) should be stationary, to avoid spurious regression results with false inference. The Augmented Dickey-Fuller test is a well-known industry standard for assessing stationarity. For this test, it is necessary to determine the number of lags to include. This is subjective; therefore, I suggest additionally analysing the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of the variables.

Secondly, the estimation window needs to be determined. Often banks measure their internal default rate data quarterly. This frequency should be matched by the MEVs that are included in the analysis. Quarterly frequency has fewer yearly observations than monthly data. The number of observations in the estimation window should be sufficient to draw valid conclusions, typically more than 30 observations. After these requirements, the LASSO BIC method can be performed. The selected MEVs are then estimated by an Ordinary Least Squares (OLS) regression. These estimated models are tested on the OLS regression assumptions: i) multicollinearity between the MEVs, ii) autocorrelation in the residuals, iii) heteroskedasticity in the residuals, and iv) normality of the residuals. Lastly, the forecast of the selected model is evaluated by an unbiasedness test, an accuracy test and an efficiency test. I conduct this evaluation to demonstrate how banks can compare the model selected by the LASSO BIC method with other models, such as those selected by their previous method. However, this step is not essential for adjusting the TtC PD to a PiT PD.

In this study I perform the above-described methodology on European corporate loan default rate data. I incorporate seven European MEVs in the LASSO BIC selection procedure: the 3M Euribor, industrial production, inflation (Harmonised Index Customer Prices), traded volume, unemployment rate, governmental bond spread, and corporate bond spread. The time series of the variables ranges from January 2005 until December 2023 with monthly frequency. I select five different models using an expanding estimation period to forecast five consecutive years separately. The training data, or the in-sample estimation windows, range from 31/01/2006 to 31/12/2022, depending on the year being forecasted, as shown in Section 3.4 **Table 1**. I transform the time series of non-stationary variables into stationary time series by applying first differencing. Additionally, I include twelve lags of these MEVs as they may hold explanatory power regarding the default rate and because monthly data is analysed (see Appendix **Table 9**). Analysing the lagged impact of macroeconomic variables on the default rate provides more insights into the relationships and dynamics at play (Bocchio et al., 2023). For quarterly data I suggest including four lags of the MEVs, as performed by Durović (2019).

The TtC PD transition probabilities are determined based on the Standardised Approach or by the Internal Rating-Based Approach. Banks do not calculate the TtC PDs directly under the Standardised Approach. Instead, they apply regulatory-set risk weights based on the asset class and the creditworthiness of the counterparty (Van Roy, 2005). Under the Internal Ratings-Based Approach banks use their own internal estimates of key risk components including PD, Loss Given Default (LGD) and Exposure at Default (EAD). The PD estimates from the IRB Approach are TtC PDs. Vasicek's One-Factor model is the industry standard for adjusting a TtC PD transition matrix to a PiT PD transition matrix. The One-Factor model computes a firm value V_i which is driven by a systematic risk factor and an idiosyncratic risk factor and by a correlation of the obligors between these two factors. After selecting the MEVs, these variables need to be individually forecasted and standardised to a common factor Z. If this realisation of the common factor is known, the individual conditional probability PD_i becomes Gaussian distributed and can be expressed in cumulative Gaussian distributions. Now the TtC PD transition probabilities can be adjusted to PiT PD transition probabilities by using Vasicek's One-factor model.

Individually forecasting the selected MEVs falls outside the scope of this research. For forecasting variables individually, the ARIMA forecasting method is a well-known and widely accepted approach in the industry. While most European banks rely on their own internal forecasts, some also incorporate forecasts provided by central banks (in Europe the EBA), or credit rating agencies (such as Moody's and S&P).

The results provide evidence of changes in the explanatory power of macroeconomic variables over time for an expanding estimation window. Although it is not practical for banks to frequently update the incorporated macroeconomic variables in their PiT PD adjustment, the selected MEVs should be re-examined whenever new observed default rate data is added. The estimated signs of these selected MEVS align with the evidence provided in several papers (Avgeri & Psillaki, 2023; Boccio et al., 2023; Nigmonov et al., 2022). Furthermore, the 10^{th} lag of the first differenced 3M Euribor rate and the 8^{th} lag of the first differenced European corporate bond spread appear to be most explanatory over the years. The finding suggests that this macroeconomic variable should be included in the selection procedure for PiT PD adjustment by European banks for corporate loan portfolios. However, few banks incorporate the European corporate bond spread as selected macroeconomic variable (MEV) for their PiT PD adjustment (EBA Monitoring Report, 2023). It depends on a bank's portfolio and internal default rate data whether the corporate bond spread is a significant explanatory variable. The forecast evaluation shows inaccuracy, mainly due to the inherent difficulty of predicting crises, which leads to more inaccurate forecasts. The LASSO BIC method is a more efficient, automated approach compared to the regression methods by Durović (2019) and Bocchio et al. (2023) and the PCA method by Breed et al. (2023).

1.1 Reading guide

This paper is structured as follows: Chapter 2 begins with a detailed stretch of the IFRS 9 regulation followed by a discussion of the existing literature on the topic in Section 2.2. Section 2.3 provides complexity theorem in financial markets and its relevance for forecasting. Chapter 3, the methodology chapter, discusses the macroeconomic variables, the PiT PD model adjustment, the LASSO BIC method, data sources, the forecast methodology, and the forecast evaluation method. Chapter 4 discusses the results and the implications for banks. Chapter 5 presents the conclusion and discusses the limitations of the study.

Chapter 2: Literature review

2.1 Regulatory details

The core business model of banks revolves around accumulating deposits from clients, constituting the bank's liabilities, and utilising these deposits for lending purposes in exchange for a certain interest rate. These loans represent the bank's assets, generating interest rate returns. Nonetheless, lending entails inherent risks, given the possibility that borrowers may default on their obligations. The Global Financial Crisis shed light on how much an economy depends on the banking system, and its soundness is heavily determined by the risk profile of lending portfolios (Bocchio, Crook & Andreeva, 2023). As a result, regulations are established to enable banks to anticipate and mitigate potential losses, ensuring they maintain adequate provisions to cover for these potential losses to maintain financial stability in the economy. On 24 July 2014 the International Accounting Standards Board (IASB) published the International Financial Reporting Standard 9 (IFRS 9), replacing the International Accounting Standard 39 (IAS 39) which was the accounting standard in place for banks to compute their required level of provisions (Commission Regulation (EU), 2016). However, this model merely used incurred losses to determine the required level of provisions. Meaning, banks only recognised credit losses when a default event had occurred to calculate the required provision. Solely past realisations were used to determine the level of the required provisions. It resulted in deferred recognition of credit losses because only the events that have occurred and current conditions influenced the credit risk evaluation. The effect of possible future credit losses was not taken into consideration in the calculations, even if they were already expected at that moment (Vasiilyeva & Frolova, 2019).

Banks were required to apply the new IFRS 9 standard for fiscal years beginning on or after 1 January 2018. The IFRS 9 framework has a forward-looking approach, compared to the prior backward-looking IAS 39. Hence, forward-looking information (FLI) is used for the computation of the required provisions of banks. Financial institutions need to consider not only past and current information, but also forward-looking components when building their impairment reports. An understanding of the dynamics of credit risk and default rates is highly important. Many papers address that the selection of the macroeconomic variables is mostly based on expert judgment by the banks incorporating this method (Basson & Van Vuuren, 2023; Vasilyeva & Frolova, 2019; Conze, 2015; EBA Monitoring report, 2023). This was also emphasised by the Basel Committee on Banking Supervision:

"A bank's use of experienced credit judgment, especially in the robust consideration of reasonable and supportable forward-looking information, including macroeconomic factors, is essential to the assessment and measurement of expected credit losses" (Basel Committee on Banking Supervision, 2015).

The IFRS 9 guidelines are internationally set out in the Basel Committee on Banking Regulation 2015. Furthermore, the IFRS9 guidelines are also set out on a European level by the European Commission of the European Union: the Commission Regulation (EU) 2016/2067. IFRS 9 only accounts for financial instruments on the balance sheet, hence financial assets and financial liabilities. Instead of an incurred model to determine the required level of the provisions, the Expected Credit Loss model is now prescribed by the IFRS 9 regulation. Banks compute their expected credit losses to determine the required amount of provisions at the current moment in time (the so-called reporting date), to account for these expected losses. Equation (1) shows the simplified version of the ECL calculation.

$$ECL = PD * LGD * EAD * DF \qquad (1)$$

Where,

ECL - Expected Credit Losses
PD - Probability of Default
LGD - Loss Given Default
EAD - Exposure at Default
DF - Discounting Factor

The Probability of Default is the possible percentage of borrowers in a lending portfolio that will default. The Loss Given Default (LGD) represents the portion of exposure that remains unrecovered in the event a borrower defaults. Even when a default occurs, there may be a recovery amount, such as collateral value in a mortgage loan, which offsets the total exposure defaulted. The Exposure at Default (EAD) quantifies the total outstanding borrowings or the maximum potential loss that the bank as a lender is exposed to when a default event occurs. Hence, the ECL is equal to the probability of default in a lending portfolio times the loss given default of this portfolio times the exposure at default times the discount factor as these possible future losses need to be discounted to the present value.

In this paper I focus on the Probability of Default model component of the ECL model, as stated in the introduction of this research. Banks need to determine which macroeconomic variables they incorporate in their adjustment of the TtC PD to a PiT PD. TtC PD estimates are calculated using long-term averages, resulting in a stable estimate that remains consistent throughout the business cycle and credit cycle (Basson & Van Vuuren, 2023). These TtC PDs are computed by either the

Standardised approach or the Internal Ratings Based approach (Basel Committee on Banking Supervision, 2005)[1]. PiT PD estimates are based on current economic conditions, incorporating all available information and forecasts. Consequently, PiT PD estimates are more reflective of real-time conditions and are significantly influenced by both the business and credit cycles (Basson & Van Vuuren, 2023; Novotny-Farkas, 2016). To adjust the TtC PD to a PiT PD, it is essential to incorporate forecasts of macroeconomic variables to compute the future probability of default transition matrix, as required by IFRS 9. This approach is logical, as numerous studies have demonstrated significant relationships between certain macroeconomic variables, such as interest rates, and corporate loan default rates (Avgeri & Psillaki, 2023; Peri & Rachedi, 2020; Ashraf & Shen, 2019).

Instead of forecasting the default rate individually, the default rate is forecasted and adjusted by macro-economic factors. Timmermann (2005) provides two main reasons for using forecast combinations: Firstly, individual forecasts may be variably impacted by structural breaks, which can arise from factors such as institutional changes or technological advancements. Certain models may adapt rapidly and experience only temporary disruptions due to structural breaks, while others possess parameters that adjust slowly to new post-break data. Secondly, individual forecasting models may be prone to misspecification bias of an unknown nature, including biases related to the underlying data-generating process and variable selection. Forecasting based on macroeconomic factors is crucial, as these have proven to be significant drivers of default risk. Incorporating the underlying relationships between default rates and the (future) state of the economy is a key feature of the IFRS 9 framework. Therefore, I propose an efficient machine learning method to select these macroeconomic variables based on their explanatory power in historical European corporate loan default rate data.

The Basel Committee on Banking Supervision (BCBS) does not prescribe a certain methodology in the selection of these macro-economic variables to incorporate forward-looking information in the PiT PD. It merely provides guidance and outlines principles and expectations:

"Consideration of forward-looking information, including macroeconomic factors, is a distinctive feature of ECL accounting frameworks and is critical to the timely recognition of ECL. Banks will have to employ sound judgment consistent with generally accepted methods for economic analysis and forecasting" (BCBS, 2015).

2.2 Existing literature

For banks within Europe there is a high variation in which macroeconomic variables they incorporate (EBA Monitoring Report, 2023). This can be due to several reasons : i) A bank's historical default rate data availability, ii) every bank has its own unique credit portfolios with different characteristics

[1] See Chapter 3.2 for the description of these approaches and the methodology to PiT adjustment

and therefore different correlations with macro-economic variables (such as business loans, mortgages, etc.) and iii) there is no unanimous method in selecting the macro-economic variables as it is not prescribed by the regulator (BCBS, 2015). As the mandatory implementation of the IFRS 9 framework is relatively new, literature is still scarce, but there are some papers discussing the selection of the macro-economic variables within the IFRS 9 framework. These consist of two papers incorporating a regression analysis (Durović, 2019; Bocchio et al. 2023) and one paper incorporating a Principal Component Analysis (PCA) (Breed et al., 2023).

Durović (2019) introduces an example of a possible retail estimation method of lifetime probability of default in accordance with IFRS 9. The paper aims to provide a robust method for estimating the PiT PD that meets the stringent requirements of IFRS 9. As required by the IFRS 9 framework, the method rests on the “term structure of probability of default” conditional to given forward-looking macroeconomic dynamics. The author uses quarterly data ranging from 2008 Q3 to 2016 Q4. To obtain the forecast of the default rate, linear regression models are estimated using 5 macroeconomic indicators as independent variables: yearly GDP growth (GDP), unemployment rate (UNEMP), quarterly consumption change (CONS), quarterly change of FX rate (FX) and quarterly inflation (INF). The paper examined all possible linear models of one, two and three independent variables combination, with time lags from one to four quarters. Models are selected that fulfil the following criteria:

1. The estimated sign of the independent variable is in line with expectation
2. All estimated coefficients are statistically significant at a 5% significance level
3. Shapiro-Wilk p-value test for residual normality is lower than 5%
4. No multicollinearity exists measured by variance inflation factor less than 5

The independent variables are modelled in the first difference due to the non-stationarity property of the observed default rate (dependent variable). Furthermore, if the model has a Durbin-Watson (DW) p-value of less than 5%, heteroskedasticity and autocorrelation consistent (HAC) standard errors are calculated, and coefficient significance is reviewed based on these standard errors.

While the method aligns with all important assumptions in time series regression analysis, it is computationally intensive and requires a lot of effort to estimate all linear regression model combinations. Secondly, the four criteria and the autocorrelation criterium need to be fulfilled per model. This requires lots of manual estimation and validation. Furthermore, this method only uses HAC standard errors when the DW-test has a p-value of less than 5%, which is a test on the autocorrelations in the residuals. However, the residuals of a model could have no significant autocorrelation while having inconsistent variance (heteroskedasticity). Therefore, it is essential to test the residuals not only for autocorrelation, but also for heteroskedasticity. Otherwise, the estimated coefficients are biased which leads to invalid inference (Stock & Watson, 2004).

Also, the Durbin-Watson test has two disadvantages. Firstly, it is difficult to interpret and secondly it only tests for autocorrelation in the first lag of the residuals of the model. The test is difficult to interpret as it does not provide a p-value but a value between 0 and 4. A value close to 2 indicates that there is no autocorrelation, whereas a value higher than 2 suggests negative autocorrelation and a value less than 2 positive autocorrelation (Savin & White, 1977). The Portmanteau's test is a more comprehensive assessment of the autocorrelation in the residuals across multiple lags and includes a p-value for significance testing. Hence, the Portmanteau's test better tests the residuals' autocorrelation structure (Stock & Watson, 2004) and is therefore used in this study.

Breed et al. (2023) propose the Principal Component Regression as methodology to adjust the IFRS 9 PD term structures for macroeconomic forecasts. The PCR is a regression technique based on the Principal Component Analysis (PCA). First, a PCA is performed on all combinations of the independent variables. Secondly, a regression is fitted on the selected principal components as covariates. The benefits of using the PCA method are that no multicollinearity exists and that the influence of variables is more evenly distributed. However, a major disadvantage of this method is that it requires to fit many model combinations for all independent variables and lags. This process is computationally intensive, particularly when dealing with many variables and possible lags. This problem is solved by implementing the LASSO BIC method, which is demonstrated in this study. This method enhances out-of-sample predictions by omitting certain regressors and by shrinking the size of the coefficients of the independent variables that have less predictive power, further explained in Section 3.3.

When constructing the principal components, PCA does not take the dependent variable into account. It solely focuses on selecting the standardised independent variables (principal components) that have the highest explanatory power of the maximum variance in the independent variables. The second step then is to scale the standardised independent variables back and perform the regression on the dependent variable, which is what makes the method the so-called PCR. For the first steps, the PCA, many combinations of the standardised independent variables and their lags need to be analysed. This is a great disadvantage, especially in comparison to the LASSO BIC machine learning technique. Furthermore, the authors refer to the criteria used for model selection as proposed by Durović (2019). However, only two criteria are used instead of the five criteria listed by Durović (2019). These two criteria are: i) the selection of models with estimated coefficient signs that align with economic expectations and ii) that these coefficients are statistically significant on a 5% level. However, it is of great importance to test the residuals of the selected models for autocorrelation and heteroskedasticity, as these are two crucial assumptions for OLS regressions. Violation of these assumptions creates bias in the estimated coefficients and therefore leads to incorrect inference (Stock & Watson, 2004).

Bocchio et al. (2023) implement a regression analysis method, which is similar to Durović's. First, the author performs a pre-assessment based on univariate analysis by estimating the default rate on one covariate at the time, considering different lag specifications for the independent variables to select a first pool of potential explanatory variables. Then several explanatory variables are combined, and the final model specification is selected such that all covariates are significant at the 10% level (so not 5% like in Durović's paper). The models are controlled for multicollinearity, heteroskedasticity and autocorrelation. If these last two assumptions are violated, they also make use of HAC standard errors. Again, this method is computationally intensive and requires performing regression estimation of many model combinations.

In comparison to these model selection methods discussed in the literature, the automated model selection method named LASSO BIC is proposed. Both the regression methods of Durović (2019) and Bocchio et al. (2023) and the PCR method of Breed et al. (2023) require evaluating numerous covariate combinations, which is highly labour-intensive. The LASSO BIC machine learning technique offers automated variable selection, meaning computational efficiency and better performance on high dimensional data (Angrist & Frandsen, 2022). The machine learning technique initially focuses on the explanatory power of the independent variables on the dependent variable and penalises for overfitting and multicollinearity by shrinking some coefficients to zero[2]. It directly incorporates the dependent variable into its selection process, aiming to improve predictive performance and interpretability. Therefore, the method is significantly more efficient and offers better interpretability than the existing methods. This leads to the following hypothesis:

Hypothesis 1: *Models selected using the Least Absolute Shrinkage and Selection Operator (LASSO) machine learning technique demonstrate the selection of macroeconomic variables that have a significant relationship with the dependent variable.*

My work differs from the literature because monthly data is used which has increased reliability, and by proposing and analysing a relatively new automated model selection method. This method could potentially replace those described in the literature due to its improved implementation efficiency. I contribute to the literature by demonstrating the implementation of the methodology and the results of the forecasts of the selected models on the European corporate loan default rate.

As a relevant addition to this research that focuses on a proposed model selection methodology, the importance of complexity theories is shortly discussed for the prediction of future credit losses and the future state of the economy.

[2] See Chapter 3.3 for a detailed description of the LASSO model selection method

2.3 Complexity theories in economics

The goal of the prescribed ECL model is to estimate the closest possible value that may occur due to the defaults of obligors on a bank's lending exposures. It is important that this estimate is as close as possible to the real future losses on a bank's lending exposures, to ensure financial stability (BCBS, 2015). The IFRS 9 regulation improves the quality and consistency of financial reporting for financial instruments (BCBS, 2015). This leads to better-informed decision-making by stakeholders and contributes to the stability and confidence of financial markets (BCBS, 2005). However, due to the uncertainty of future events it is impossible to have predictions that are in line with reality; there's fundamental uncertainty for models trying to predict the future. This is especially the case for unlikely events that have a large impact, such as the recent Covid-19 pandemic or the war in Ukraine. These events are not predictable through traditional statistical or inductive methods, making it challenging to learn from them effectively (Taleb, 2005). This is also proven by the number of overlays incorporated by many European banks upon their calculated ECL/provision level (EBA Monitoring Report, 2023). Furthermore, besides this description of uncertainty in exogenous events occurring, there are underlying processes in financial markets which cause unpredictability. While future crises are unpredictable, I do include a dummy for the Global Financial Crisis to account for the increased default rate values, thereby enhancing the reliability of my models.

Analysts often make use of Gaussian statistics (normal distributions) in their models/calculations. However, financial markets are not normally distributed, proven by realised data where the tail of losses is much bigger than according to a normal distribution (Taleb, 2005; Dacorogna & Pictet, 1997; Sowdagur & Narsoo, 2017). Additionally, rather than making rational and optimised decisions, people engage in satisficing; they do not have access to all information and base their choices on the options they know (Simon, 1956). Furthermore, individuals exhibit bounded rationality and behavioural biases; our cognitive processing capacity is limited, leading us to make decisions influenced by these cognitive biases (Simon, 1956; Gigerenzer & Gaissmaier, 2011). The current IFRS 9 framework tries to predict the expected credit losses as accurate as possible. The complexity theories are not incorporated into the ECL model prescribed by the IFRS 9 regulation, and therefore not in the selected LASSO BIC models of this research. I shortly point out these theories, to describe the possible underlying reasoning of having ECL estimates that do not align with reality. Given that the objective of the IFRS 9 regulation is to estimate Expected Credit Losses with high accuracy, it is important not to overlook these theories, as the actual characteristics of financial markets often differ from those assumed by credit models.

Chapter 3: Methodology and Data

3.1 Macro-economic determinants of default rates

This section discusses several forward-looking macroeconomic variables that potentially have explanatory power over historical corporate loan default rates. These include the factors: the Euribor 3M interest rate, Industrial Production (as proxy for GDP), Inflation (HICP), Trading volume (XVOL), Unemployment rate, Governmental Bond Spread and Corporate Bond Spread.

3.1.1 Euribor 3M interest rate

The 3M Euribor, or 3-Month Euro Interbank Offered Rate, is a key interest rate benchmark in the Eurozone that reflects the average rate at which European banks lend money between each other over a three-month period. This European interbank interest rate is widely used in financial markets for pricing various financial instruments, including corporate loans, bonds and other derivatives (Hazenberger & Pfarrhofer, 2021; Vergote & Gutiéterrez; Andersen & Wagener, 2002). Several studies have shown a significant impact of increasing interest rates on the corporate loan default rates (Avgeri & Psillaki, 2023; Peri & Rachedi, 2020; Ashraf & Shen, 2019). An increase in interest rates leads to an increase in the probability of loan defaults (Nigmonov et al., 2022; Ashraf & Shen, 2019). Therefore, there might be a significant relation between interest rates and corporate loan defaults. The following hypothesis is tested:

Hypothesis 2: *An increase in the 3M Euribor rate has a significant positive relation with the observed default rates of corporate loans in Europe.*

3.1.2 European Industrial Production (GDP)

GDP growth is often used as one of the macroeconomic variables to adjust the TtC PD to a PiT PD (EBA Monitoring Report, 2023). However, GDP growth is measured with quarterly frequency. This may be sufficient if the institution’s historical default rate data is also measured with quarterly frequency. However, in this paper, monthly data is analysed. Also, as a common tool to increase the data frequency, institutions have implemented data interpolation on the quarterly GDP growth data. However, this method could raise measurement errors and biases in the analysis. Hence, as performed in many other studies that face the problem of having monthly frequent data, Industrial Production was used as a proxy for GDP growth (Ascari & Haber, 2022; Bocchio, Crook & Andreeva, 2023; Beaudreau, 2005). This variable is included in the analysis because positive changes indicate economic growth and higher incomes, which may significantly relate to the default rate of European corporate loans. An increase of the Industrial Production possible reduces transitions of corporate loans towards delinquency (Bocchio et al., 2023). Several studies have proven a long-run negative relationship between GDP growth and loan defaults, indicating that an increase in GDP may lead to fewer loan defaults (Bocchio et al., 2023; Fatouh & Giansante, 2022; Alsamara et al., 2019). Therefore, the following hypothesis is tested:

Hypothesis 3: *An increase in the Industrial Production as proxy for the GDP has a significant negative relation with the observed default rates of corporate loans in Europe.*

3.1.3 European Harmonised Index Consumer Prices (HICP)

The inflation rate affects the European economy by reducing the purchasing power of both consumers and companies, thereby decreasing economic growth. Rising inflation devalues the real worth of money, resulting in diminished purchasing power and lower real returns on investments (Eldomiaty et al., 2020). One of the tools employed by the European Central Bank (ECB) to account for the moving inflation, is the adjustment of interest rates. This measure serves to incentivise market participants, while also influencing the dynamics of costs and profitability. Furthermore, an increase in interest rates encourages individuals to allocate more funds towards savings, given the enhanced returns available through deposit accounts. Conversely, lower interest rates prompt participants to increase borrowing and spending, thereby stimulating consumption and overall economic activity. Inflation is a good indicator of the economic status of Europe for different types of assets (Marshall, 1992). An increase in the inflation rate leads to lower spending on goods and services. This impact can directly affect corporations’ revenue and profit margins, particularly for consumer-facing industries, making it harder for these businesses to meet their loan obligations. Studies have shown a positive relation between a country’s or region’s inflation and the loan defaults (Nigmonov et al., 2022; Durović, 2019). Therefore, the following fourth hypothesis is formulated:

Hypothesis 4: *An increase in the inflation has a significant positive relation with the observed default rates of corporate loans in Europe.*

3.1.4 European Traded Volume (XVOL)

The traded volume on an international level can be an indicator of the economic health of a certain region (Capannelli et al., 2009). A decline in traded volume could signal a weakening economy, which might lead to reduced consumer spending and lower corporate revenues, especially for companies that rely heavily on international markets. A decrease in exports, for instance, possibly leads to reduced sales and profits for exporting firms, affecting their ability to service debt and increasing the likelihood of defaults. Furthermore, the traded volume in European markets can reflect liquidity conditions. Lower liquidity means it is harder for companies to refinance their debts by issuing new bonds or stocks, increasing their default risk. This can lead to financial difficulties and higher default risks (Longstaff & Schaefer, 2014). Therefore, the following fifth hypothesis is formulated:

Hypothesis 5: *An increase in the traded volume has a significant negative relation with the observed default rates of corporate loans in Europe.*

3.1.5 European Unemployment rate

The unemployment rate is measured as the percentage of unemployed people in the labour force. The labour force includes individuals that are of working age, typically between 15 and 65 years old, which depends on the specific country, and who are either employed or actively searching for employment (Maas, 2020). An increase in the unemployment rates often signals a deterioration of the economic health. It possibly leads to decreased consumer spending as, on an overall level, consumers have less disposable income. This decrease in consumer spending can further impact corporate revenues and overall economic growth (Gruber, 1997). This could potentially lead to more companies facing financial difficulties, which increases the corporate loan default risk. Therefore, the following sixth hypothesis is formulated:

Hypothesis 6: *An increase in the unemployment rate has a significant positive relation with the observed default rates of corporate loans in Europe.*

3.1.6 European Governmental Bond Spread

There are several papers that have shown evidence that the governmental yield spread is a good predictor of future economic activity (Leombroni et al., 2021; Lee, 2021; Evgenidis et al., 2020). The European governmental bond spread illustrates the yield on European government bonds across various maturities. In this paper, the European governmental bond spread is computed by subtracting the European 1-year maturity bond yield from the 30-year maturity bond yield, which is a common methodology (Giesecke et al., 2011; Driessen, 2003). An upward sloping yield curve indicates that the yields on long-term bonds are higher than those on short-term bonds, forming what is typically referred to as a normal or positive yield curve (de Lint & Stolin, 2003). A decreasing slope can be a sign of a future recession. The slope of the bond yield curve is measured by the yield spread (Leombroni et al., 2021). As this yield spread possibly signals the future state of the economy, it could have significant explanatory power on the European corporate loan default rate (Bleaney & Veleanu, 2021). Therefore, the following seventh hypothesis is formulated:

Hypothesis 7: *An increase in the governmental bond spread has a significant negative relation with the observed default rates of corporate loans in Europe.*

3.1.7 European Corporate Bond Spread

The corporate bond spread is determined by measuring the difference in yield between a corporate bond and a comparable government bond (ICE Data Indices, 2022). "Comparable" in this context denotes that the spread is calculated using government bonds that share identical characteristics in terms of coupon rates, maturity dates, amortisation, and call schedules. Giesecke et al. (2011) state the following in their working paper about the relation between corporate credit spreads and the probability of default:

"Because of the link between credit spreads and the intensity of the underlying process triggering default, corporate credit spreads contain information about the probability of default. Specifically, credit spreads can be expressed in terms of the actual probabilities of default plus a component representing a risk premium compensating investors for bearing credit risk" (Giesecke et al., 2011).

Hence, the corporate bond spread is a measurement for the additional compensation investors demand for taking on the higher risk. The higher risk is often due to deteriorating financial conditions of the issuing companies, leading to an increased probability of default. Several papers have shown a significant relationship between corporate bond spreads and the probability of default (Merton, 1974; Elton et al., 1999; Giesecke et al., 2011; Driessen, 2003). Therefore, the eighth hypothesis is formulated:

Hypothesis 8: *An increase in the corporate bond spread has a significant positive relation with the observed default rates of corporate loans in Europe.*

3.2 IFRS 9 Point-in-Time methodology

In general, the ECL should be calculated by the equation shown before (1) in Section 2.1. However, the challenging part of the IFRS 9 framework is in the situation that the credit quality of a loan deteriorates. It is necessary for banks to distinct its financial instruments in three credit risk stages since initial recognition:

Stage 1: **Performing stage.** Financial instruments which showed no significant increase of credit risk since initial recognition (SICR); provisions equal to 12-months' ECL (or for the whole life if it is less than 12 months). The interest return is calculated based on the balance sheet value/gross carrying amount.

Stage 2: **Underperforming stage.** Financial instruments that have shown a significant increase in credit risk (SICR) since initial recognition, but are not impaired, require provisions equal to the expected credit loss (ECL) for the entire lifetime of the asset. The interest return is calculated as the effective interest on balance sheet value/gross carrying amount.

Stage 3: **Non-performing stage.** Credit impaired financial instruments are assigned to this stage; provisions equal to the ECL for the whole lifetime of an asset. Interest revenue is calculated as the effective interest on amortised cost carrying amount (i.e. net of credit allowance).

Hence, depending on the stage, the Expected Credit Losses (considering future states) of banks need to be calculated by the ECL model, as prescribed by the IFRS 9 framework. The portfolio ECL is the aggregate ECL over all loans. For stage 1 the formula is the following:

$$ECL_{Stage\ 1} = \frac{(\sum_{n=1}^{N}\sum_{t=1}^{T} PD_{t,n}^{PiT} * LGD_{t,n}^{PiT} * EAD_{t,n}^{PiT})}{1+EIR^{t}} \quad (2)$$

Where,

${PD^{PiT}}_{t,n}$ - Probability of Default of loan *n* between time *t-1* and *t*

${LGD^{PiT}}_{t,n}$ - Loss Given Default of loan *n* between time *t-1* and *t*

${EAD^{PiT}}_{t,n}$ - Exposure At Default of loan *n* between time *t-1* and *t*

t - period (year) of calculation of ECL

T - remaining time until loan maturity, which for stage 1 is 12 months (one year)

N - total number of loans within the portfolio

r is the effective interest rate as discount factor (discounting for 12 months for stage 1), which for stage 1 is equal to effective interest rate on the balance sheet/gross carrying amount

The formula for calculating the ECL for stage 2 and stage 3 is similar, however the ECL under stage 1 only needs to be computed for a maturity of one year, while for stage 2 and 3 it is computed for the remaining time until loan maturity. The ECL for stage 3 is computed by the summation of the LGD times the summation of the EAD divided by the effective interest rate for discounting. The PD component is equal to one for stage 3, as the PD is 100% due to the occurrence of default. Furthermore, for stage 3, the ECL is discounted by the effective interest rate on amortised cost instead of the effective interest rate on the gross carrying amount. The provisions are calculated by the following formula:

$$Provisions = ECL_{Stage\ 1} + ECL_{Stage\ 2} + ECL_{Stage\ 2} \quad (3)$$

Where,

$ECL_{Stage\ 1}$ is the provision of financial instruments assigned to Stage 1

$ECL_{Stage\ 2}$ is the provision of financial instruments assigned to Stage 2

$ECL_{Stage\ 3}$ is the provision of financial instruments assigned to Stage 3

Measuring the ECL has become the basis for determining the amount that a financial institution must hold to act as a buffer to protect against potential impairments (BCBS, 2015). All three ECL model components (PD, LGD, EAD) should reflect relevant prevailing or future macroeconomic scenarios. Due to the complexity and the size of the ECL model, this paper focuses on the Probability of Default model component of the ECL model. While stage 2 and 3 need calculation of the ECL for the lifetime of the financial instruments, the same selection procedure of the macroeconomic variables (LASSO BIC), which is proposed in this paper, can be performed on the other two stages. The focus of this paper is to propose an efficient selection procedure of the most explanatory macroeconomic variables based on historical portfolio default rate data, while forecasting these macroeconomic variables separately is the next step in the procedure to compute a Point-in-Time Probability of Default. Moreover, generally, financial institutions make use of internal macroeconomic forecasts, even if quite a large share of institutions have also relied on central banks' forecasts, at least for the baseline scenario (EBA Monitoring Report, 2023). The selection of the most explanatory macroeconomic variables on the default rates of a specific financial institution is one step in the process of rescaling the TtC PD to a PiT PD. The rescaling of the TtC PD to a PiT PD consists of several stages, which is further explained in this section.

The credit losses of financial institutions can be divided in two parts: i) the Expected Credit Losses and ii) the Unexpected Credit Losses (UCL). Basel II regulation focuses on the UCL. This is the regulatory capital requirement for banks to absorb potential unexpected losses. To be able to absorb these unexpected losses banks need to have enough capital. Financial institutions can choose between two main approaches to calculate credit risk under the regulatory requirements: the Standardised Approach and the Internal Ratings-Based (IRB) Approach as prescribed in Basel II by the BCBS (2005).

Standardised approach

Under the Standardised Approach banks use fixed risk weights for assets, which are determined by external credit assessments (like ratings from credit rating agencies) or prescribed by regulation for unrated exposures (Konno & Itoh, 2016). Banks do not calculate PDs directly under the Standardised Approach. Instead, they apply regulatory-set risk weights based on the asset class and the creditworthiness of the counterparty (Van Roy, 2005). Based on this TtC PD transition matrix the PiT PD transition matrix needs to be estimated under IFRS 9.

Internal Ratings-Based (IRB) Approach

Under the Internal Ratings-Based Approach banks use their own internal estimates of key risk components including PD, LGD and EAD. The PD estimates from the IRB Approach are TtC PDs. TtC PDs, which are non-procyclical, are used for regulatory credit risk capital requirements. TtC PDs reflect the average likelihood of default over a full economic cycle, smoothing out the fluctuations due to economic conditions (Liu, 2017). Financial institutions estimate a TtC transition matrix (usually 1-year) based on historical data that records migrations between internal rating grades, including the state of default. TtC PDs are directly calculated in the IRB Approach. This is advantageous for implementing the prescribed IFRS 9 regulation, as the TtC PDs are already computed. Only an adjustment of this ECL model component to a PiT PD is necessary to reflect the relevant prevailing or future macroeconomic scenarios, rather than the long-term average (Basson & Van Vuuren, 2023). Under IFRS 9, banks must recognise an impairment loss based on expected credit losses rather than incurred losses. The calculated ECL is equal to the required level of provisions. Hence, for PiT PD modelling, the TtC PD computed in the IRB Approach or in the Standardised Approach are used and adjusted by incorporating forward looking information (FLI).

Next, the explanation of converting the TtC PD transition matrix to a PiT PD transition matrix is provided. This process begins with an understanding of the Merton model and then delves into Vasicek's One-Factor Firm's model.

Merton's model

The mapping function used to derive the conditional PiT PDs from the average TtC PDs is an adaption of Merton's (1974) single asset model to credit portfolios (BCBS, 2005). Merton's model was an early adopted method to model corporate defaults. It is an influential benchmark model that uses a firm value approach to determine default:

$$V_t = S_t + F_t \quad (4)$$

Where,

V_t - Stochastic asset value

S_t - Equity value

F_t - Debt value

The Merton model assumes that the asset value (V_t) follows a log-normal distribution. The debt of a firm consists of a zero-coupon bond with a face value F and maturity T. A default occurs when the firm is not able to repay the debt at time T, hence, when the asset value of the firm is lower than the debt value. This leads to two scenarios:

$$V_T >= F,\ S_T >= V_T - F \quad (5)$$

$$V_T < F,\ F_T = V_T \ and \ S_T = 0 \quad (6)$$

In scenario one the asset value is equal or bigger than the debt value (5). Hence, the shareholders receive the residual value of the asset value minus the debt value. In the second scenario the asset/firm value is less than the debt value (6). The shareholders hand over control to the bond holders and they liquidate the firm. The debt holders receive the asset value (6). This is smaller than F, but they receive what is still left of the asset value. Therefore, F is equal to V and the shareholders receive nothing (Merton, 1974). This results in the following maximum equations for the two scenarios:

$$S_T = [V_T - F]^+ \tag{7}$$

$$F_T = F - [F - V_T]^+ \tag{8}$$

The shareholders receive a maximum equation (7), and the debt holders receive a maximum equation (8). The maximum value of a shareholder is unlimited but could also become zero, whereas the value of a bond holder is either F or lower when the company defaults, so in the situation of the second scenario (6). In Merton's model the equity of a firm (S_T) is a call option on the asset value (V_T) of a firm with strike price F. The bond holders receive a fixed amount, but are short in a put option (Merton, 1974). The Merton model is not sophisticated enough as there exist dependencies between defaults. Hence, the Merton model was extended to the One Factor model to account for default dependencies (Fermanian, 2020).

One-Factor Firm's value model

To account for default dependencies between obligors the One-Factor model was introduced, also named the One Factor Firm's value model, defined by Vasicek (1987) and Belkin et al. (1998). This firm value V_i is driven by a systematic risk factor and an idiosyncratic risk factor and by its correlation between these two factors:

$$V_i(T) = \sqrt{\rho_i}Y + \sqrt{1-\rho_i}\epsilon_i\ , i = 1, \ldots, N. \tag{9}$$

Where,

$V_i(T)$ - Asset value of the *i*-th obligor at time *t*, $V_i \sim\ N(0,1)$

Y - Common factor, $Y \sim\ N(0,1)$

ϵ_i - idiosyncratic factors, $\epsilon_i \sim\ N(0,1)$

Obligor *i* defaults if $V_i(T) \leq d_i$

Asset values $V_1, \ldots, V_N$ of different obligors are correlated with covariance matrix.

Given the default probabilities, the proper default thresholds can be calibrated. This can be computed by the following:

$$P(V_i(T) \leq d_i) = \phi(d_i) = p_i \Leftrightarrow d_i = \phi^{-1}(p_i) \tag{10}$$

To calibrate the default threshold, an asset value from a standard normal distribution that is equal or lower than the default threshold is needed, which must be equal to the probability of default. The default threshold can then be determined by computing the inverse of the cumulative distribution function (CDF). The computation of the default threshold is not fully relevant for this paper.

However, it is important to understand the One-Factor model, which is explained further in this chapter.

A key assumption of the Gaussian One-factor model is that an obliger i's default event is driven by a standard normal systematic factor and a standard normal obliger-specific factor (Belkin et al., 1998). Another assumption is credit risk homogeneity in the portfolio (Vasicek, 1987), meaning that individual loans or credit exposures within a portfolio share similar or identical risk characteristics. When this common factor (Y) is realised, the defaults become independent as they are now only modelled by idiosyncratic factors. Based on the realisation of the common factor, the individual conditional default probability $PD_i(y)$ can be computed:

$$PD_i(y) = P(V_i(T) < d_i | Y = y) \tag{11}$$

$$= P(\sqrt{\rho_i}\, Y + \sqrt{1-\rho_i}\, \epsilon_i < d \mid Y = y)$$

$$= P(\epsilon_i < \frac{d_i - \sqrt{\rho_i}\, Y}{\sqrt{1-\rho_i}} \mid Y = y)$$

$$= \phi(\frac{d_i - \sqrt{\rho_i}\, y}{\sqrt{1-\rho_i}}) = \phi(\frac{\phi^{-1}(PD_i) - \sqrt{\rho}\, Y}{\sqrt{1-\rho_i}})$$

If the realisation of Y is known, the individual conditional default probability PD_i is Gaussian distributed and can be expressed in cumulative gaussian distributions. This results in the well-known Vasicek's (1987) formula used by Basel II regulation, where the realised common factor is noted as Z (BCBS, 2006):

$$PD_i(Z) = \phi(\frac{\phi^{-1}(PD_i) - \sqrt{\rho}\, Z}{\sqrt{1-\rho_i}}) \tag{12}$$

Where,

ϕ - Cumulative distribution function (CDF) of a normal (Gaussian) distribution

PD_i - Individual conditional default probability

The conditional probability of default can be used as a proxy of the future default rate driven by the unknown systematic factor on a large homogeneous portfolio, i.e. on a portfolio with many exposures with the same unconditional PD and the same correlation parameter ρ (Basson & Van Vuuren, 2023). A macro-economic model (linear regression) predicting the state of the economy (Z) is developed based on internal data. It is important to understand this formula in the context of PiT PD modelling under IFRS 9. The reason being, that this model is used to adjust the TtC PD transition matrix to a PiT PD transition matrix. The future state of the economy is computed by standardising the forecasted macroeconomic variables that were selected.

It can be written as the following equation to adjust the TtC PDs to PiT PDs:

$$PD^{PiT}{}_i(Z_t) = \phi\left(\frac{\phi^{-1}(PD^{TTC}{}_i) - \sqrt{\rho}\, Z_t}{\sqrt{1-\rho_i}}\right) \tag{13}$$

Z_t is a representation of the status of the economic conditionals as well as thereafter (Basson & Van Vuuren, 2023). By incorporating the standardised forecasted state of the economy in this formula the TtC PD transition matrix is adjusted to a PiT transition matrix.

To summarise, Merton's model is used as benchmark model for default modelling. However, the Merton model does not account for default dependencies between the obligors. Therefore, this model was extended to the One Factor model by Vasicek (1987) to account for correlation (dependency) between the obligors. This model is used in the IFRS 9 framework to adjust the TtC PD transition matrix to a PiT PD transition matrix. The Z-model includes forecasted—not observed at the reporting date—values of macro-economic variables. To make the PD model forward looking, it must be made explanatory by incorporating observations and forecasts of macroeconomic factors. This involves calculating probabilities of default conditional on these factors (Conze, 2015). This vector of macroeconomic variables then needs to be normalised to one realised common factor (Z_t). Based on the forecasted Z score, the TtC PD migration probabilities are adjusted using the Vasicek model to compute the PiT PD transition matrix. Hence, based on Vasicek's application, a financial institution needs to determine the most explanatory macroeconomic variables of its historical default rate data, forecast these individually, then standardise this to a realised common factor (Z_t) and eventually compute the PiT PD transition matrix.

3.3 LASSO model selection methodology

The Least Absolute Shrinkage and Selection Operator (LASSO) machine learning technique, introduced by Tibshirani (1996), represents a type of penalised regression. Estimators that use LASSO solely for variable selection are said to be "post-lasso" according to Angrist and Frandsen (2022). Hence, the application in this paper is the "post-lasso" method as it is only used for model selection. Instead of a manual and time-consuming regression analysis, analysts can use the LASSO method as an automated selection method of independent variables in an OLS regression (Belloni et al., 2014). It enhances out-of-sample predictions by omitting certain regressors and by shrinking the size of the coefficients of the independent variables that have less predictive power. When variables are correlated, LASSO selects the variable with the greatest explanatory power for the dependent variable and shrinks the coefficients to zero of the other correlated variables, thus effectively handling multicollinearity.

The LASSO technique offers two possible implementation approaches: cross-validation (CV) or Bayesian Information Criterion (BIC). CV randomly samples the data with each run, whereas the

Bayesian Information Criterion (BIC) approach utilises a fixed estimation period without random sampling (Bellone et al., 2014). Consequently, the LASSO BIC method consistently produces the same results when applied to identical data sets, making it simpler to interpret and execute. Furthermore, the BIC method ensures model parsimony by balancing the model's complexity (the number of additional variables) against its fit to the data. BIC achieves this by penalising the likelihood function based on the number of parameters in the model and the sample size, which tends to favour simpler models when comparing models that fit the data equally well (Spanos, 2010; Stock & Watson, 2002). Additional independent variables should have enough explanatory power of the dependent variable relative to the penalty of adding an additional variable, as this increases the complexity of the model. The LASSO BIC serves as a model selection method that ensures parsimonious model selection and tackles the problem of model overfitting. By assessing the BIC values of different models, the model presenting the lowest BIC is identified as the best fitting model.

3.3 Data Sources

In this paper the European corporate loan default rate was analysed from 01/31/2005 to 31/12/2023. As financial institutions' corporate loan default rate data is highly sensitive and proprietary, European default rate data was analysed on an aggregated scale. These data could be extracted from Moody's Data Library (2024). Moody's is a globally recognised credit rating agency that provides comprehensive financial data, analytical tools, credit ratings, and research. It is one of the major credit rating agencies alongside Standard & Poor's and Fitch Ratings. The data has monthly frequency, resulting in a total amount of 228 observations shown in the Appendix **Table 9** named 'Defaultrates'. A high frequency and therefore a large sample size is important for a valid and high-quality time series analysis (Stock & Watson, 2004; Szigel & Gyürüs, 2023). Reliability of tests on time series, such as stationarity tests, increase with the sample size. In banking practice, it is a widespread approach to "enlarge" the sample sizes of default rate time series by increasing the observation frequency. This is often done from annual to quarterly frequency, but monthly frequency increases the sample size even more and therefore the reliability of the prediction (Szigel & Gyürüs, 2023). Therefore, in this paper, I analyse monthly frequent data. However, it is important to note that banks may typically have access to data merely on a quarterly basis. While valid inference can be drawn based on quarterly data, it is crucial to consider the sample size.

The monthly data for various European macroeconomic variables were mostly sourced from the European Central Bank's Data Portal (2024). This dataset includes the unemployment rate, trade volume, inflation, the 3-month Euribor interest rate, and government bond spread. The European corporate bond spread was imported from the Federal Reserve Bank of St. Louis (FRED). The industrial production time series data was imported from Eurostat (2024). The descriptive statistics of all these macroeconomic variables are also provided in Appendix **Table 9**.

I include a dummy variable for the Global Financial Crisis period to the analysis. The reason being, that from the period of around 2008 to 2010 there is a big spike in the European corporate loan default rates. After considering of the distribution of the default rate data, I concluded that this spike comprises vertical outliers in the OLS regression. While vertical outliers are acceptable in regression analysis, a dummy was included to account for it (Stock & Watson, 2004). The dummy is named GFC, as shown in Appendix **Table 9.**

3.3.1 Data transformations

To perform a valid LASSO BIC model selection and regression forecast, both the dependent variable and the independent variable should be stationary (Stock & Watson, 2004). This is required to avoid spurious regressions (Szigel & Gyürüs, 2023). Therefore, I perform Augmented Dicky-Fuller tests (ADF) on the variables to test for stationarity and, additionally, analyse the ACF and PACF plots. In addition to performing the ADF test, I recommend analysing the ACF and PACF plots, as the determination of the number of lags for the ADF test is subjective and requires manual input. I transformed the default rate time series data to stationary data by first differencing, which is often used to make time series data stationary (Stock & Watson, 2004). This time series is visualised in the Appendix **Figure 2.** The same holds for all independent variables that prove to be non-stationary by the ADF test, which are also visualised in the Appendix **Figure 3-9**. Furthermore, lagged versions of the independent variables were included in the analysis, as they may hold explanatory power regarding the dependent variable. Analysing the lagged impact of macroeconomic variables on the default rate can provide more insights into the relationships and dynamics at play (Bocchio et al., 2023).

The data of all variables consist of 228 observations with dates ranging from 31/01/2005 – 31/12/2023 with monthly frequency. The first difference is computed by subtracting the value at *t* from the value prior to *t*, so at *t-1*. Therefore, after taking the first difference the value at 31/01/2005 cannot be computed and is excluded, leading to a total of 227 observations. Taking the lagged version of a differenced variable leads to the exclusion of the value at the next timestep leading to 226 observations, up until a total of 216 observations for the 12^{th} lagged version of a variable. Hence, the LASSO BIC analysis starts from 31/01/2006 until 31/12/2022 as shown in **Table 1** below**.**

3.4 Forecasting methodology

The training data, or the in-sample estimation periods, range from 31/01/2006 to 31/12/2022, depending on the year being forecasted. An overview of the estimation periods is given in the following table:

Table 1. *In-sample estimation periods of corporate loan default rates*

	In-sample period	In-sample period	In-sample period	In-sample period	In-sample period
From	31/01/2006	31/01/2006	31/01/2006	31/01/2006	31/01/2006
To	31/12/2018	31/12/2019	31/12/2020	31/12/2021	31/12/2022
Selection method	LASSO BIC	LASSO BIC	LASSO BIC	LASSO BIC	LASSO BIC
Prediction year	2019	2020	2021	2022	2023

I perform the LASSO model selection method on the training data/in-sample period. After selection of the model, it is estimated on the in-sample period to forecast out-of-sample by an Ordinary Least Squared (OLS) regression. For each out-of-sample forecast, the in-sample estimation period increases one year starting with the forecast of 2019, and ending with the largest in-sample estimation period for the prediction of 2023. As the in-sample estimation period changes for each prediction, there is a possibility that the selection of certain macroeconomic variables changes as their explanatory power may change over time. Therefore, the following hypothesis is formulated:

Hypothesis 9: *As the in-sample estimation period extends annually from 2019 to 2023, the set of macroeconomic variables selected by the LASSO model varies, reflecting changes in their explanatory power over time.*

After first having performed the LASSO BIC model selection and then model estimation, several assumptions need to be tested to draw valid inference and conclusions of these predictions. Therefore, I perform the following steps:

1. The regressors/independent variables do not exhibit multicollinearity. As the LASSO selection procedure handles multicollinearity, it still is tested for consistency. The Variance Inflation Factor (VIF) test is used, which is a diagnostic tool to detect multicollinearity.
2. The residuals of the model do not contain autocorrelation. The Portmanteau's test is used and recommended as it provides a p-value and has better interpretability than the Durbin-Watson test.[3]
3. The residuals of the model do not contain heteroskedasticity. To test the validity of this assumption, the Breusch-Pagan/Cook-Weisberg test was performed.
4. Lastly, the residuals of the model are normally distributed. The Shapiro-Wilk test on normality was performed to test this assumption.

[3] See Chapter 2.2 for detailed reasoning

The VIF for a predictor X_i is calculated as:

$$VIF_i = \frac{1}{1-R_i^2} \tag{14}$$

Where,

R_i^2 – The coefficient of determination from a regression X_i on all other predictors.

The Null hypothesis of the test is that there is no multicollinearity. When the VIF value is 1 or close to 1 there is no multicollinearity, when the value is between 1 and 5 there is moderate multicollinearity (which is generally accepted) and bigger of equal to five means there is high collinearity.

The Portmanteau's test for autocorrelation, or also known as the Box-Pierce or Ljung-Box test, computes the following test statistic:

$$Q = n(n+2)\sum_{k=1}^{m}\frac{\hat{\rho}_k^2}{n-k} \tag{15}$$

Where,

n – sample size

$\hat{\rho}_k^2$ – autocorrelation at lag k

m – number of lags being tested

The Null hypothesis of the test is that there is no autocorrelation up to lag m, hence that the residuals are independently distributed. I chose five lags for testing the residuals on autocorrelation, which is a common choice. Testing on a higher number of lags possibly leads to more model complexity, which is not desired as a parsimonious model is preferred. Testing the autocorrelation of the residuals up to five lags provides a balance between capturing sufficient historical information and maintaining model simplicity to avoid overfitting. The significance of the test statistic is used to either accept or reject the null hypothesis.

The Breusch-Pagan/Cook-Weisberg test for heteroskedasticity involves regressing an auxiliary regression of the squared residuals from the original regression model. The test statistic is then given by:

$$X^2 = nR^2 \tag{16}$$

Where,

n – number of observations

R^2 – coefficient of determination from the auxiliary regression

The Null hypothesis is that the residuals are homoscedastic, hence that the residuals contain a constant variance. The significance of the test statistic is used to either accept or reject the null hypothesis.

The Shapiro-Wilk test for normality computes the following test statistic:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (17)$$

Where,

$x_{(i)}$ – the ordered sample values

$x_{(i)}$ – the sample mean

a_i – constants generated from the means, variances, and covariances of the order statistics of a sample size *n* from a normal distribution

The Null hypothesis is that the residuals of the model are normally distributed. The significance of the test statistic is used to either accept or reject the null hypothesis.

If the multicollinearity assumption is violated, which is unlikely as LASSO BIC handles multicollinearity, the most explanatory variable among the correlated variables is selected by LASSO BIC. Violation of the autocorrelation or/and the heteroskedasticity assumption is handled by using Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors before drawing conclusions on the significance of the coefficients. Furthermore, if the normality assumption is violated this is handled by relying on the Central Limit Theorem (CLT). CLT states that, given a sufficiently large sample size, the sampling distribution of the sample mean (or sum) will tend to be normally distributed. Specifically, the large sample size ensures that the estimators of the regression coefficients are asymptotically normal, which justifies the use of standard statistical tests and confidence intervals despite the non-normality of the residuals.

3.5 Forecast evaluation

Three tests are performed to evaluate the predictions of the selected models:

1. The Unbiasedness test
2. The Accuracy test
3. The Efficiency test

Unbiasedness test

This test is performed to test whether a model is systematically overestimating or underestimating the actual values. An unbiased model has predictions that are, on average, correct, meaning that the forecast errors are random and not systematic. Hence, whether the average forecast error significantly differs from 0 by using a t-test on the following equation:

$$\mathbb{E}\left[\varepsilon_{t++1|t}\right] = \mathbb{E}\left[y_{t+1} - y_{t+1|t}\right] = 0 \qquad (18)$$

The Null hypothesis of the test is that the forecast is unbiased. The significance of the t-statistic is considered to either accept or reject the null hypothesis.

Accuracy test

The Mean Squared Prediction Error (MSPE) metric is used to test the accuracy of the forecasts. It tests the accuracy by squaring the forecast errors and by dividing these by the number of observations. The computation of the MSPE is given in the following equation:

$$MSPE = \frac{1}{P}\sum_{t=T_1}^{T-1}(y_{t+1} - F_{t+1|t})^2 \quad (19)$$

Where,

P = $T - T_1$

y_{t+1} = true value

$F_{t+1|t}$ = forecasted value conditional on *t*

The MSPE metric is sensitive to large errors due to the squaring of the forecast errors, making it a robust measure of prediction accuracy. A low MSPE value indicates a more accurate forecast. However, interpreting the MSPE is difficult, because it depends on the scale of the time series. Therefore, when evaluating the accuracy of a single model using MSPE, a better approach is to compare the MSPE value to the out-of-sample variance. A lower MSPE compared to the out-of-sample variance indicates that the model provides reliable and accurate forecasts.

Efficiency test

It should not be possible to forecast the forecast error using information that is available at time *t*. This is tested by the efficiency test. A regression is used where the forecasted value is regressed on the actual value with a constant. The forecast is efficient if the true value is equal to the forecasted value. Therefore, the coefficient of the forecasted value in the regression should be equal to 1 and the expected error term should be equal to 0 (the constant), which is tested by a joint F-test. The following regression is performed for the efficiency test:

$$y_{t+1} = \beta_0 + \beta_1 F_{t+1|t} + \eta_{t+1} \quad (20)$$

Where,

$F_{t+1|t}$ = the forecasted value

η_{t+1} = error term

Where it is tested that $\beta_0 = 0$ and $\beta_1 = 1$ with a joint F-test. Hence, the Null hypothesis is:

H_0: $\beta_0 = 0\ and\ \beta_1 = 1$. The significance of the F-statistic is considered to either accept or reject the null hypothesis.

Chapter 4: Results

In this section the results are shown and discussed. First, the models selected by LASSO BIC are provided, followed by the regression results. These results, along with the significance of the estimated coefficients, are used to answer the hypotheses outlined in Chapter 2 and 3. The differences between the forecasts and the actual achieved returns are visualised and discussed. Furthermore, the forecast errors are computed to evaluate the forecast on Unbiasedness, Accuracy and Efficiency. This evaluation determines the quality of the forecast and discusses the implications of the results for the IFRS 9 framework.

4.1 The LASSO BIC selected models

The following table shows the models selected by LASSO BIC per forecasted year. Five years are forecasted with an expanding in-sample training dataset, as explained in Section 3.4.

Table 2. *Selected LASSO BIC models per forecasted year*

Variables	Forecast 2019	Forecast 2020	Forecast 2021	Forecast 2022	Forecast 2023
Lag(D(Euribor), -10)	x	x	x	x	x
Lag(Spread Corp.), -8)		x	x	x	x
Lag(IP, -5)	x				
D(Spread Gov.)					x
D(Spread Corp.)					x
Lag(D(HICP), -11)					x
Lag(D((Spread Gov), -1)					x
Lag(D(Spread Corp.), -11)					x
GFC					x
_cons	x	x	x	x	x

Legend:
x - estimated

Hence, the following five OLS regression models (I,II,III,IV and V) are selected by LASSO BIC to forecast the European Corporate loan default rate:

Regression I (forecast 2019): $Default\ Rate_t = \alpha + \beta_1(Euribor_{D10})_t + \beta_2(IP_{L5})_t + \varepsilon_t$

Regression II (forecast 2020): $Default\ Rate_t = \alpha + \beta_1(Euribor_{D10})_t + \beta_2(Spread\ Corp_{D8})_t + \varepsilon_t$

Regression III (forecast 2021): $Default\ Rate_t = \alpha + \beta_1(Euribor_{D10})_t + \beta_2(Spread\ Corp_{D8})_t + \varepsilon_t$

Regression IV (forecast 2022): $Default\ Rate_t = \alpha + \beta_1(Euribor_{D10})_t + \beta_2(Spread\ Corp_{D8})_t + \varepsilon_t$

Regression V (forecast 2023): $Default\ Rate_t = \alpha + \beta_1(Euribor_{D10})_t + \beta_2(Spread\ Corp_{D8})_t +$
$\beta_3(Spread\ Gov_D)_t + \beta_4(Spread\ Corp_D)_t + \beta_5(HICP_{D11})_t + \beta_6(Spread\ Gov_{D1})_t +$
$\beta_7(Spread\ Corp_{D11})_t + \beta_8 GFC_t + \varepsilon_t$

Where,

$Default\ Rate_t$ - the forecasted European corporate loan default rate at time *t*

α - the constant of the regression

$(Euribor_{D10})_t$ - the 10th lag of the first differenced 3M Euribor at time *t*

$(IP_{L5})_t$ - the 5th lag of the MoM European Industrial Production

$(Spread\ Corp_{D8})_t$ - the 8th lag of the first differenced European Corporate Bond Spread at time *t*

$(Spread\ Gov_D)_t$ - the first differenced European Governmental Bond Spread at time *t*

$(Spread\ Corp_D)_t$ - the first differenced European Corporate Bond Spread at time *t*

$(HICP_{D11})_t$ - the 11th lag of the first differenced European Harmonised Index Consumer Prices at time *t*

$(Spread\ Gov_{D1})_t$ - the 1st lag of the first differenced European Governmental Bond Spread at time *t*

$(Spread\ Corp_{D11})_t$ - the 11th lag of the first differenced European Corporate Bond Spread at time *t*

GFC_t - the Global Financial Crisis dummy at time *t*

ε_t - the error term

Each regression model is estimated using Ordinary Least Squares (OLS). However, as explained in Section 3.4, after each OLS estimation a model needs to be tested on the OLS assumptions. The results on the VIF multicollinearity test for each model is around one (see Appendix **Table 10**), meaning that all five models do not contain multicollinearity, which is in line with expectation as LASSO handles multicollinearity. The Shapiro-Wilk test on the normality of the residuals suggests for all five models that the residuals are not normally distributed as the p-value of all five tests are significant at a 1% level. Hence, the Null hypothesis, stating that the residuals are normally distributed, is rejected for all five models. However, this violation is handled by CLT due to the large sample sizes, as explained in Section 3.4.

For the second and third assumptions, tests need to be performed on the autocorrelation and on heteroskedasticity in the residuals. The model residuals of the 2019 forecast show no significant autocorrelation left and no significant heteroskedasticity, as shown in the Appendix **Table 11**. Therefore, standard errors are used to determine the significance of the coefficient estimates, shown in **Table 3** below.

Table 3. OLS regression results 2019-2020

Variables	Forecast 2019
Lag(D(Euribor), -10)	1.92003***
	(0.39438)
Lag(IP, -5)	-0.25345***
	(0.05491)
Constant	0.00050
	(0.00062)
Observations	156
Adj. R-squared	0.2483

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The Adjusted R-squared of the model is around 25%, meaning that 25% of the variation of the first difference European corporate loan default rate is explained by the model. This is an acceptable adjusted R-squared as a very high adjusted R-squared might suggest overfitting, meaning that the model is optimised on the in-sample data but does not perform well out-of-sample. The in-sample estimation window ranges from 31/01/2006 to 31/12/2018 with monthly frequency, which has a total of 156 observations. In this model the 10^{th} lag of the first differenced 3M Euribor rate is included and the 5th lag of the MoM Industrial Production growth rate. The 3M Euribor rate has an estimated positive coefficient of around 1.92 that is statistically significant at the 1% level, with a p-value < 0.01. The 5^{th} lag of the MoM Industrial Production growth rate has an estimated negative coefficient of around -0.25 that is also statistically significant at the 1% level, with a p-value < 0.01.

The residuals of the 2020 forecast model contain no significant autocorrelation (p-value of 0.1248) but do contain significant heteroskedasticity (p-value of 0.04) on a 5% level (see Appendix **Table 11**). Furthermore, the residuals of the 2021, 2022 forecast models also do not contain significant autocorrelation but do contain significant heteroskedasticity. The residuals of the 2023 forecast model contain both autocorrelation and heteroskedasticity (see Appendix **Table 11**). Therefore, HAC standard errors were employed to ensure valid interpretation of the significance of the coefficients across these models. The regression results are shown in **Table 4** below:

Table 4. OLS regression results 2021-2023

Variables	Forecast 2020	Forecast 2021	Forecast 2022	Forecast 2023
D(Spread Gov.)				0.58813**
				(0.23673)
D(Spread Corp.)				0.15396**
				(0.06122)
Lag(D(Euribor), -10)	1.85539***	1.88386***	1.86773***	1.66552***
	(0.49487)	(0.49855)	(0.49196)	(0.60631)
Lag(D(Spread Corp.), -8)	0.21258**	0.19461**	0.19798**	0.18429**
	(0.10267)	(0.09358)	(0.09213)	(0.08026)
Lag(D(HICP, -11)				0.32944**
				(0.15450)
Lag(D(Spread Gov. -1)				0.54149*
				(0.27772)
Lag(D(Spread Corp., -11)				-0.15034*
				(0.08971)
GFC				0.00505**
				(0.00245)
Constant	0.00035	0.00048	0.00031	0.00014
	(0.00060)	(0.00056)	(0.00055)	(0.00048)
Observations	168	180	192	204
Adj. R-squared	0.1908	0.1843	0.1842	0.2978

HAC Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The Adjusted R-squared of the models are around 20% except for the 2023 model that has an adjusted R-squared around 30%. These are again acceptable results and in line with the industry such as the results of Durović (2019). The observations of each regression make a jump of twelve, which is in line with the methodology, because for each forecast, the estimation window expands one year. In the 2020, 2021 and 2022 forecast models, the same two independent variables are selected: the 10th lag of the first differenced 3M Euribor rate and the 8th lag of the first differenced corporate bond spread. The estimated coefficients of the 3M Euribor are significantly positive and around 1.86 for these three models at a 1% level. The estimated coefficients of the corporate bond spread are around 0.2 and significantly positive but at a 5% level.

The 2023 forecast model contains several independent variables: The first differenced governmental bond spread, the first differenced corporate bond spread, the 10th lag of the first differenced 3M Euribor, the 8th lag of the first differenced corporate bond spread, the 11th lag of the first differenced HICP, the 1st lag of the first differenced governmental bond spread, the 11th lag of the first differenced governmental bond spread and the GFC dummy. The estimated coefficient of the 3M Euribor is significantly positive and around 1.67 at a 1% level. Three versions of the corporate bond spread are included: the first difference, the 8th lag of first difference and the 11th lag of the first difference. The first two are positive and significant at a 5% level. The 11th lag of the first difference corporate bond spread is negative but only significant at a 10% level. Two versions of the governmental bond spread are included: the first difference and the 1st lag of the first difference. The first difference is positive and significant at a 5% level while the 1st lagged version is positive but significant at a 10% level. The 11th lag of the first differenced European Harmonised Index Consumer Prices has a significantly positive effect on the dependent variable at a 5% level. The Global Financial Crisis (GFC) dummy is significantly positive with a coefficient close to zero at a 5% level.

4.2 Conclusions on hypotheses

The first hypothesis, formulated in Section 2.2, states that the models selected by LASSO demonstrate macroeconomic variables that have a significant relationship with the European corporate loan default rate. This hypothesis is accepted based on the regression results shown in **Table 3** and in **Table 4** in the previous Section. Most selected macroeconomic variables show a significant relation with the European corporate loan default rate at a 5% level. Among all five models, only the forecast for the year 2023 shows that the 1st lag of the first differenced government bond spread and the 11th lag of the first differenced corporate bond spread are insignificant at the 5% level.

In the methodology chapter 3.1 several hypotheses were formulated about the possible relationships between the macroeconomic variables and the European corporate loan default rate. The OLS regression results are used to derive conclusions regarding these hypotheses.

The second hypothesis is accepted, which states that there is a significant positive relation between the 3M Euribor and the European Corporate Loan default rate. The regression results support this, as for all five models the estimated coefficient for the 10th lag of the first differenced 3M Euribor is significantly positive across all models. This demonstrates that there is a lagged positive relationship between the 3M Euribor and the dependent variable at a 1% confidence level. Meaning, an increase in the 10th lagged interest rate results in an increase in the European corporate loan default rate.

The third hypothesis is accepted as for the 2019 forecast model the 5th lag of the MoM Industrial Production was selected, which has a significant negative relation with European corporate loan default rate at a 1% confidence level. This means that an increase in the 5th lagged industrial production results in a decrease in the European corporate loan default rate for the 2019 forecast model.

The fourth hypothesis is also accepted as the model of the 2023 forecast provides evidence of a significant positive relationship between the 11th lag of the first differenced HICP and the European corporate loan default rate at a 5% confidence level. Hence, the evidence supports that an increase in the 11th lagged European inflation results in an increase in the European corporate loan default rate.

The fifth hypothesis is not accepted. The traded volume variable was not selected by the LASSO BIC method, indicating that it does not have sufficient explanatory power and is not statistically significant in explaining the default rates of corporate loans in Europe. Hence, there is no evidence to either support or reject the fourth hypothesis.

The sixth hypothesis is not accepted. The same reasoning applies as for the fifth hypothesis; the variable was not selected, hence there is no evidence.

The seventh hypothesis is accepted, as the model of the 2023 forecast provides evidence of a significant positive relationship between the 10th lag of the first differenced European governmental bond spread and the European corporate loan default rate at a 5% confidence level. Hence, an increase in the 10th lag of the governmental bond spread, which means a future upstate of the economy, results in a decrease in the European corporate default rate.

The eighth hypothesis is accepted, as the model of the 2023 forecast provides evidence of a significant positive relationship between the first differenced and the 8th lag of the first differenced European corporate bond spread and the European corporate loan default rate at a 5% confidence level. Hence, an increase in the European corporate bond spread, which suggests that investors want a higher compensation for the credit default risk, results in an increase in the European corporate loan default rate.

In addition to the hypotheses regarding potential relationships between macroeconomic variables and the European corporate loan default rate, another hypothesis was formulated. This nineth hypothesis states that the selected macroeconomic variables vary over time due to the expanding estimation window, suggesting that the explanatory power of some macroeconomic variables may increase or decrease over time. The LASSO BIC model selection results indicate that the selected macroeconomic variables indeed change over time. Therefore, the nineth hypothesis is accepted. This is an important takeaway for banks that utilise specific selected variables for their PiT PD adjustments. Over time, the explanatory power of these variables may vary, with some becoming more significant while others may diminish in relevance. However, it is often impractical for banks to frequently update their selected variables for PiT adjustments. Nonetheless, it is important to be aware of these potential changes and consider them in the computation of their Expected Credit Losses.

4.3 Forecast results

The OLS estimations are used to forecast the European corporate loan default rate. **Table 5** below shows the descriptive statistics of the actual out-of-sample default rates compared to the forecasted default rates and of the forecast error.

Table 5. Descriptive statistics actual default rates, forecasted default rates and forecast error

Variable	Obs	Mean	Std. Dev.	Min	Max
Actual Default Rate	60	.035	.021	.011	.078
Forecasted Default Rate	60	.001	.003	-.003	.01
Forecast Error	60	0	.01	-.048	.044

The average of the actual default is 0.035 and the average of the forecasted default rate is 0.001, indicating a deviation between the two. Furthermore, regarding the standard deviation, the minimum value and the maximum value indicate a notable deviation between the actual and the forecasted default rates. These observations are supported by **Figure 1**:

Figure 1. Actual Default Rate plotted against the predicted Default Rate (2019-2023)

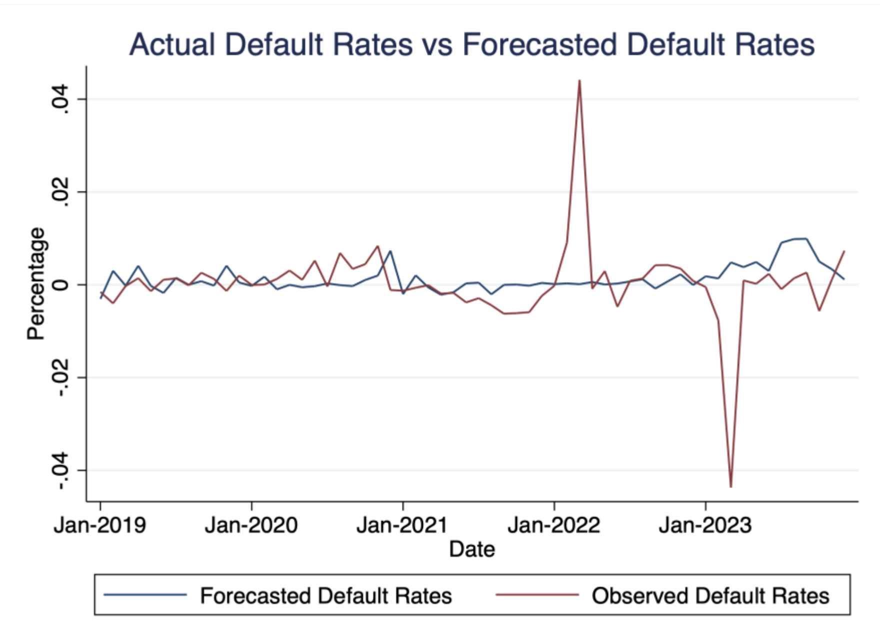

Figure 1 shows that on average the forecasted default rate follows the actual defaults rate relatively close, except for two big spikes in the observed default rate in the beginning of 2022 and the beginning of 2023. In February 2022 Russia invaded Ukraine, leading to sharp increases in energy prices which led to a substantial economic disruption across Europe. Furthermore, many companies still were financially struggling in the aftermath of the Covid-19 pandemic of prolonged knockdowns and reduced consumer demands. These factors possibly drove the high spikes in European corporate loan default rates. As discussed before, crises are very difficult to predict. Therefore, within IFRS 9 and the prescribed ECL method, the regulator should consider complexity theories and scenario thinking for an improved credit risk framework. It is for these unpredictable crises that banks also incorporate manual adjustments/overlays to their estimated provisions based on the ECL, as banks, models are not able to capture these sudden events with their models (EBA Monitoring Report, 2023). This also confirms the problem raised regarding forecasting by complexity theories, stating that financial markets contain fundamental uncertainty, especially proven for crisis periods.

Three tests were performed to evaluate the quality of the European corporate loan default rate forecast. The Unbiasedness test shows no significant constant, as shown in **Table 6**. Therefore, the Null hypothesis is accepted. This means that the model is not systematically overestimating or underestimating the actual values; the predictions are, on average, correct, meaning that the forecast errors are random and not systematic. The accuracy test suggests that the forecast is not accurate, because the out-of-sample variance is smaller than the computed MSPE. The discrepancy is primarily due to the forecast's inability to capture the spikes caused by the discussed crises. Furthermore, the forecast is not efficient as the joint F-test shows significance, with a p-value of 0.0040 as shown in

Table *8*, hence the Null hypothesis is rejected. While the forecast is not accurate, the LASSO BIC method does select variables that provide a dynamic forecast, where the inaccuracy is mainly driven by the crisis periods, which are difficult to predict. In practice these unpredictable events are handled by manual adjustments of the calculated ECL level.

Table 6. Unbiasedness test

Variable	Coefficient	T-statistic	P-value
Constant	-0.0009678	-0.78	0.436
	(0.0012)		

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7. Accuracy test

Out-of-sample variance	0.0000777
MSPE	0.0000908
Difference	-0.0000132

Table 8. Efficiency test

Hypothesis	F(2, 58)	Prob > F
H0: $\beta_0 = 0 \; and \; \beta_1 = 1$	6.07	0.0040***

*** p<0.01, ** p<0.05, * p<0.1

The forecast evaluation tests show evidence of an unbiased forecast, but also evidence of an inaccurate forecast. However, this does not violate the outcome of the research question, as the LASSO BIC method does provide an efficient method of selecting explanatory variables and the quality of the forecast is driven by several factors, such as omitted variable bias. As the goal is to determine the most explanatory variables for the IFRS 9 framework, the method does show to efficiently select significant variables as shown in the regression results in **Table 3** and **Table 4** in Section 4.1.

In this research I perform the LASSO BIC method to capture the most explanatory macroeconomic variables to adjust the TtC PD to a PiT PD. These MEVs are then used to perform a forecast of the European corporate loan default rate, to show the extent to which these explain the default rate and contain forecast power. This serves as an evaluation method to test the forecast power of the models containing the MEVs and determine the final MEVs for the PiT PD adjustment. The selected MEVs are not changed frequently, as frequent adjustments are impractical for banks. For the PiT PD adjustment these variables need to be individually forecasted, for example by using ARIMA forecasting, which has shown high forecasting quality. This was out of the scope of this research. Banks also choose to utilise forecasts provided by central banks (in Europe the EBA) or credit rating agencies (such as Moody's and S&P). The high reliance on internal projections has caused some variability across institutions in the forecasted macroeconomic variables figures embedded in the ECL models, which has naturally resulted in divergent effects of the FLI incorporation and final ECL model's outputs (EBA monitoring report, 2023). The forecasted macroeconomic variables are then combined into a vector and normalised to a single common factor (Z_t). This normalised factor is used to compute the PiT PD transition probabilities using Vasicek's equation (13) in Section 3.2.

Chapter 5: Conclusion

This paper aims to propose the LASSO BIC machine learning method for selecting macroeconomic variables (MEVs) for the PiT PD adjustment within the IFRS 9 framework. By analysing European corporate loan default rate data, I demonstrate that the LASSO BIC method efficiently selects macroeconomic variables that have significant explanatory power of the European corporate loan default rate. Significant relationships have been identified between certain macroeconomic variables and the European corporate loan default rate, with the direction of the relationship aligning with the evidence presented in several studies. Furthermore, the 10^{th} lag of the first differenced 3M Euribor rate and the 8^{th} lag of the first differenced European corporate bond spread appear to be most explanatory over the years. The finding suggests that this macroeconomic variable should be included in the selection procedure for PiT PD adjustment by European banks for corporate loan portfolios. The results also provide evidence of changes in the explanatory power of macroeconomic variables over time for an expanding estimation window. Although it is not practical for banks to frequently update the incorporated macroeconomic variables in their PiT PD adjustment, the selected MEVs should be re-examined whenever new observed default rate data is added. The LASSO method is a more efficient, automated approach compared to the regression methods by Durović (2019) and Bocchio et al. (2023) and the PCA method by Breed et al. (2023).

Evaluation of the out-of-sample performance is demonstrated as a valuable process for banks to test the performance of their models incorporating macroeconomic variables. However, for IFRS 9, it is only necessary to identify the most explanatory variables, and this prediction step is not required. A subsequent step would be to forecast the selected macroeconomic variables individually, which was out of the scope of this research. Based on Vasicek's application, a financial institution needs to determine the most explanatory macroeconomic variables of its historical default rate data, forecast these individually, then standardise this to a realised common factor (Z_t) and eventually calculate the PiT PD transition matrix by using Vasicek's equation (13), derived from the One-factor model.

Additionally, this study highlights the importance of considering complexity theories and diverse modelling approaches to better understand and mitigate credit risk in an uncertain economic environment. This study provides evidence of a fundamental unpredictability of financial markets, especially for crisis periods. Overall, the application of the LASSO machine learning technique presents an efficient solution for MEV selection, contributing to the implementation of the IFRS 9 framework in European banks.

5.1 Limitations

In this research I included seven MEVs in the LASSO BIC selection procedure on the European corporate loan default rate data. The results are limited to the MEVs included in the selection procedure. There may be omitted variables that have explanatory power on the default rate. However, the variables included were selected based on the existing literature. Other variables were initially included in the analysis, such as the Euronext100 returns, the cost of borrowing for corporations, the European volatility index (VSTOXX), the US/EUR exchange rate, and the 1-year government bond yield. However, after examining the correlations among all variables, these were excluded. I suggest that banks determine which MEVs to include based on the characteristics of their lending portfolios. For instance, in this study, I analysed the default rate of an aggregated European corporate loan portfolio, making variables like the 3M Euribor relevant. However, for a mortgage portfolio, it would be more relevant to include variables such as the housing price index.

As Durović (2019) states, the selection of any proposed modeling technique is highly dependent on data availability. This study was conducted using long-period monthly data, but many institutions may have access to much less data. Additionally, the choice of the number of lags incorporated in the analysis of the MEVs affect the results. For quarterly data I suggest including four lags, like Durović (2019).

The historical default rate data of a bank is highly dependent on a bank's specific Definition of Default (DoD), which can vary between institutions. While regulators provide guidelines and standards for defining default, the exact criteria can be tailored and interpreted differently by each bank.

Forecasting the MEVs themselves is outside the scope of this research. Banks often use forecasts provided by organisations like Moody's, S&P or central banks. Future research could focus on implementing the LASSO BIC method and secondly forecast these selected MEVs individually to finally compute and demonstrate the PiT transition probabilities. Furthermore, the implementation of other machine learning methods, such as neural network analysis or the random forest model, could be explored on this subject.

Finally, while this study uses an expanding estimation window, further research could consider using a rolling estimation window, which banks might find favourable as it gives more weight to recent observations.

References

Alsamara, M., Mrabet, Z., Jarallah, S., & Barkat, K. (2019).
The switching impact of financial stability and economic growth in Qatar: Evidence from an oil-rich country. *The Quarterly Review of Economics and Finance*, *73*, 205-216.

Andersen, A. B. & Wagener, T. (2002).
Extracting Risk Neutral Probability Densities by Fitting Implied Volatility Smiles: Some Methodological Points and an Application to the 3m Euribor Futures Options Prices. Available at SSRN: https://ssrn.com/abstract=359060

Angrist, J. D., & Frandsen, B. (2022).
Machine labour. *Journal of Labour Economics*, *40*(S1), S97-S140

Arthur, W.B. (2013).
Complexity Economics: A Different Framework for Economic Thought. Oxford University Press, 2013.

Ascari, G., & Haber, T. (2022).
Non-linearities, state-dependent prices and the transmission mechanism of monetary policy. *The Economic Journal*, *132*(641), 37-57.

Ashraf, B. N., & Shen, Y. (2019).
Economic policy uncertainty and banks' loan pricing. *Journal of Financial Stability*, *44*, 100695.

Avgeri, E., & Psillaki, M. (2023).
Factors determining default in P2P lending. *Journal of Economic Studies*.

Basel Committee on Banking Supervision. (2015).
Guidance on credit risk and accounting for expected credit losses.

Basson, L. J., & Van Vuuren, G. (2023).
Through-the-cycle to Point-in-time Probabilities of Default Conversion: Inconsistencies in the Vasicek Approach. *International Journal of Economics and Financial Issues*, *13*(6), 42.

BCBS. (2005).
An Explanatory Note on the Basel II IRB Risk Weight Functions. Switzerland: BIS.

BCBS. (2006).
Basel II: Part 2: The First Pillar-Minimum Capital Requirements. Switserland: BIS

Beaudreau, B. C. (2005).
Engineering and economic growth. *Structural Change and Economic Dynamics*, *16*(2), 211-220.

Belloni, A., Chernozhukov, V., & Hansen, C. (2014).
Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, *81*(2), 608-650.

Bleaney, M., & Veleanu, V. (2021).

Redenomination risk in eurozone corporate bond spreads. *The European journal of finance*, *27*(13), 1303-1325.

Bocchio, C., Crook, J., & Andreeva, G. (2023).

The impact of macroeconomic scenarios on recurrent delinquency: A stress testing framework of multi-state models for mortgages. *International Journal of Forecasting*, *39*(4), 1655-1677.

Bordo, M. D., Duca, J. V., & Koch, C. (2016).

Economic policy uncertainty and the credit channel: Aggregate and bank level US evidence over several decades. *Journal of Financial stability*, *26*, 90-106.

Breed, D. G., Hurter, J., Marimo, M., Raletjene, M., Raubenheimer, H., Tomar, V., & Verster, T. (2023).

A Forward-Looking IFRS 9 Methodology, Focussing on the Incorporation of Macroeconomic and Macroprudential Information into Expected Credit Loss Calculation. *Risks*, *11*(3), 59.

Bruche, M., & González-Aguado, C. (2010).

Recovery rates, default probabilities, and the credit cycle. *Journal of Banking & Finance*, *34*(4), 754-764.

Capannelli, G., Lee, J. W., & Petri, P. A. (2009).

Developing indicators for regional economic integration and cooperation. *Singapore Economic Review, Forthcoming, Asian Development Bank Regional Economic Integration, Working Paper*, (33).

Commission Regulation (EU) 2016/2067 (2016).

Amending Regulation (EC) No 1126/2008 adopting certain international accounting standards in accordance with Regulation (EC) No 1606/2002 of the European Parliament and of the Council as regards International Financial Reporting Standard 9.

Conze, A. (2015).

Probabilities of default for impairment under ifrs 9. *Available at SSRN 2685099*

Dacorogna, M.M. and Pictet, O.V., Heavy Tails in High-Frequency Financial Data (1997).

Available at SSRN: https://ssrn.com/abstract=939 or http://dx.doi.org/10.2139/ssrn.939

Deyoung, R., Gron, A., Torna, G., & Winton, A. (2015).

Risk overhang and loan portfolio decisions: Small business loan supply before and during the financial crisis. *The Journal of Finance*, *70*(6), 2451-2488.

Diamond, D. W., Kashyap, A. K., & Rajan, R. G. (2017).

Banking and the evolving objectives of bank regulation. *Journal of Political Economy*, *125*(6), 1812-1825.

Driessen, J. (2005).

Is default event risk priced in corporate bonds? *The Review of Financial Studies*, *18*(1), 165-195.

Đurović, A. (2019).
Macroeconomic approach to point in time probability of default modeling–IFRS 9 challenges. *Journal of Central Banking Theory and Practice*, *8*(1), 209-223.

EBA monitoring report. (2023).
IFRS 9 implementation by EU institutions. https://extranet.eba.europa.eu/sites/default/documents/files/document_library/Publications/Reports/2023/1063709/Final%20Report%20on%20IFRS 9%20implementation%20by%20EU%20institutions.pdf?retry=1

Eldomiaty, T., Saeed, Y., Hammam, R., & AboulSoud, S. (2020).
The associations between stock prices, inflation rates, interest rates are still persistent: Empirical evidence from stock duration model. *Journal of Economics, Finance and Administrative Science*, *25*(49), 149-161.

Elton, E. J., Gruber, M. J., Agrawal, D., & Mann, C. (2001).
Explaining the rate spread on corporate bonds. *The Journal of Finance*, *56*(1), 247-277.

Emmer, S., & Tasche, D. (2004).
Calculating credit risk capital charges with the one-factor model. *The Journal of Risk*, *7*(2),
1.European Central Bank – Data Portal. (2024). https://data.ecb.europa.eu/

Eurostat - Database (2024). https://ec.europa.eu/eurostat/data/database

Evgenidis, A., Papadamou, S., & Siriopoulos, C. (2020).
The yield spread's ability to forecast economic activity: What have we learned after 30 years of studies?. *Journal of Business Research*, *106*, 221-232.

Fatouh, M., & Giansante, S. (2020).
Expected Loss Model and the Cyclicality of Bank Credit Losses and Capital Ratios. Available at SSRN: https://ssrn.com/abstract=3728699

Fermanian, J.D. (2020).
On the Dependence Between Default Risk and Recovery Rates in Structural Models. Available at SSRN: https://ssrn.com/abstract=3678549

Giesecke, K., Longstaff, F. A., Schaefer, S., & Strebulaev, I. (2011).
Corporate bond default risk: A 150-year perspective. *Journal of financial Economics*, *102*(2), 233-250.

Giesecke, K., Longstaff, F. A., Schaefer, S., & Strebulaev, I. A. (2014).
Macroeconomic effects of corporate default crisis: A long-term perspective. *Journal of Financial Economics*, *111*(2), 297-310.

Gigerenzer, G., & Gaissmaier, W. (2011).
Heuristic decision making. *Annual review of psychology*, *62*, 451-482.

Gruber, J. (1997).

The Consumption Smoothing Benefits of Unemployment Insurance, The American Economic Review 87, 192.

Hauzenberger, N., Pfarrhofer, M., & Stelzer, A. (2021).
On the effectiveness of the European Central Bank's conventional and unconventional policies under uncertainty. *Journal of Economic Behavior & Organization*, *191*, 822-845.

ICE Data Indices. (2022).
Bond Index methodologies.
https://www.ice.com/publicdocs/data/Bond_Index_Methodologies.pdf

Konno, Y., & Itoh, Y. (2016).
An alternative to the standardised approach for assessing credit risk under the Basel Accords. *Cogent Economics & Finance*, *4*(1), 1220119.

Lee, U. (2021).
Another Look at the Predictive Power of the Yield Spread: New Evidence. *Journal of Accounting and Finance*, *21*(2).

de Lint, C. R., & Stolin, D. (2003).
The predictive power of the yield curve: a theoretical assessment. *Journal of Monetary Economics*, *50*(7), 1603-1622.

Liu, Y. (2017).
Bayesian Analysis of Uncertainty in Internal Rating Based PD Models. *Available at SSRN 2968708*.

Maas, B. (2020).
Short-term forecasting of the US unemployment rate. *Journal of Forecasting*, *39*(3), 394-411.

Merton, R.C. (1974).
On the pricing of corporate debt: The risk structure of interest rates. Journal of Finance 29, 449 - 470.

Miu, P., & Ozdemir, B. (2005).
Practical and theoretical challenges in validating Basel parameters: key learnings from the experience of a Canadian bank. *Journal of Credit Risk*, *1*(4), 89-136.

Moody's – Data Library. (2024).
https://www.moodys.com/

Neisen, M., & Schulte-Mattler, H. (2021).
The effectiveness of IFRS 9 transitional provisions in limiting the potential impact of COVID-19 on banks. *Journal of Banking Regulation*, *22*(4), 342-351.

Nigmonov, A., Shams, S., & Alam, K. (2022).
Macroeconomic determinants of loan defaults: Evidence from the US peer-to-peer lending market. *Research in International Business and Finance*, *59*, 101516.

Novotny-Farkas, Z. (2016).

The interaction of the IFRS 9 expected loss approach with supervisory rules and implications for financial stability. *Accounting in Europe*, *13*(2), 197-227.

Peri, A., & Rachedi, O. (2020).
Financial development, default rates and credit spreads. *The Economic Journal*, *130*(626), 534-553.

Phelan, G. (2017).
Correlated default and financial intermediation. *The Journal of Finance*, *72*(3), 1253-1284.

Savin, N. E., & White, K. J. (1977).
The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors. *Econometrica*, *45*(8), 1989–1996. https://doi.org/10.2307/1914122

Schoemaker, P. J. (1991).
When and how to use scenario planning: a heuristic approach with illustration. *Journal of forecasting*, *10*(6), 549-564.

Simon, H. A. (1956).
Rational choice and the structure of the environment. *Psychological review*, *63*(2), 129.

Sowdagur, V., & Narsoo, J. (2017).
Forecasting value at risk using GARCH and extreme value theory approaches for daily returns. *International Journal of Statistics and Applications*, *7*(2), 137-151.

Spanos, A. (2010).
Akaike-type criteria and the reliability of inference: Model selection versus statistical model specification. *Journal of Econometrics*, *158*(2), 204-220.

Stepankova, B., & Teply, P. (2023).
Consistency of banks' internal probability of default estimates: Empirical evidence from the COVID-19 crisis. *Journal of Banking & Finance*, *154*, 106969.

Stock, J. H., & Watson, M. W. (2002).
Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, *97*(460), 1167-1179.

Taleb, N. (2005).
The black swan: Why don't we learn that we don't learn. *NY: Random House, 1145*.

Tibshirani, R. (1996).
Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267-288.

Timmermann, A. (2005).
Forecast Combinations (November 2005). CEPR Discussion Paper No. 5361, Available at SSRN: https://ssrn.com/abstract=878546

Van Roy, P. (2005).
Credit ratings and the standardised approach to credit risk in Basel II.

Vasilyeva, A. & Frolova, E.A. (2019).
Methods of Calculation of Expected Credit Losses Under Requirements of IFRS 9. *Journal of Corporate Finance Research*, Vol. 13, No. 4, pp. 74-86 (2019), Available at SSRN: https://ssrn.com/abstract=3638946

Vergote, O., & Gutiérrez, J. M. P. (2012).
Interest rate expectations and uncertainty during ECB governing council days: evidence from intraday implied densities of 3-month Euribor. *Journal of Banking & Finance*, *36*(10), 2804-2823.

Appendices

Table 9. *Descriptive statistics Default rate and macro-economic variables 31/01/2005 to 31/12/2023*

Variable	Obs	Mean	Std. Dev.	Min	Max
Defaultrates (in %)	228	.032	.023	0	.132
3M Euribor (in %)	228	.011	.017	-.006	.051
IP growth (MoM % change)	228	.001	.022	-.187	.137
HICP (in %)	228	.021	.021	-.006	.106
XVOL growth (MoM % change)	227	.004	.027	-.15	.082
Unemployment rate (in %)	228	.091	.017	.065	.122
Bond Spread Gov (in %)	228	.016	.01	-.01	.037
Corp Bond spread (in %)	228	.053	.034	.019	.22
GFC (dummy)	228	.079	.271	0	1
D(Defaultrates)[4]	227	0	.009	-.063	.074
D(Euribor)	227	0	.002	-.009	.006
D(HICP)	227	0	.004	-.016	.015
D(Unemployment rate)	227	0	.001	-.002	.005
D(Bond Spread Gov)	227	0	.002	-.006	.01
D(Corp Bond spread)	227	0	.009	-.031	.071
Lag(D(Euribor),-1)[5]	226	0	.002	-.009	.006
Lag(D(Euribor),-2)	225	0	.002	-.009	.006
Lag(D(Euribor),-3)	224	0	.002	-.009	.006
Lag(D(Euribor),-4)	223	0	.002	-.009	.006
Lag(D(Euribor),-5)	222	0	.002	-.009	.006
Lag(D(Euribor),-6)	221	0	.002	-.009	.006
Lag(D(Euribor),-7)	220	0	.002	-.009	.006
Lag(D(Euribor),-8)	219	0	.002	-.009	.006
Lag(D(Euribor),-9)	218	0	.002	-.009	.006
Lag(D(Euribor),-10)	217	0	.002	-.009	.006
Lag(D(Euribor),-11)	216	0	.002	-.009	.006
Lag(D(Euribor),-12)	215	0	.002	-.009	.006
Lag(IP, -1)	227	.001	.022	-.187	.137
Lag(IP, -2)	226	0	.022	-.187	.137
Lag(IP, -3)	225	.001	.023	-.187	.137
Lag(IP, -4)	224	.001	.023	-.187	.137
Lag(IP, -5)	223	.001	.023	-.187	.137
Lag(IP, -6)	222	.001	.023	-.187	.137
Lag(IP, -7)	221	.001	.023	-.187	.137
Lag(IP, -8)	220	.001	.023	-.187	.137
Lag(IP, -9)	219	.001	.023	-.187	.137
Lag(IP, -10)	218	.001	.023	-.187	.137
Lag(IP, -11)	217	.001	.023	-.187	.137
Lag(IP, -12)	216	.001	.022	-.187	.137
Lag(D(HICP), -1)	226	0	.004	-.016	.015
Lag(D(HICP), -2)	225	0	.004	-.016	.015
Lag(D(HICP), -3)	224	0	.004	-.016	.015
Lag(D(HICP), -4)	223	0	.004	-.016	.015
Lag(D(HICP), -5)	222	0	.004	-.016	.015
Lag(D(HICP), -6)	221	0	.004	-.016	.015

[4] D(Defaultrates) means the first differenced 'Defaultrates' data

[5] Lag(D(Euribor), -1) means the first lag of the first differenced 3M Euribor data. Per independent variable lags are added to the analysis up until the 12th lagged version of the variable.

Lag(D(HICP), -7)	220	0	.004	-.016	.015
Lag(D(HICP), -8)	219	0	.003	-.016	.015
Lag(D(HICP), -9)	218	0	.004	-.016	.015
Lag(D(HICP), -10)	217	0	.003	-.011	.015
Lag(D(HICP), -11)	216	0	.003	-.011	.015
Lag(D(HICP), -12)	215	0	.003	-.011	.015
Lag(XVOL, -1)	226	.004	.027	-.15	.082
Lag(XVOL, -2)	225	.004	.027	-.15	.082
Lag(XVOL, -3)	224	.004	.027	-.15	.082
Lag(XVOL, -4)	223	.004	.027	-.15	.082
Lag(XVOL, -5)	222	.004	.027	-.15	.082
Lag(XVOL, -6)	221	.004	.027	-.15	.082
Lag(XVOL, -7)	220	.005	.027	-.15	.082
Lag(XVOL, -8)	219	.005	.027	-.15	.082
Lag(XVOL, -9)	218	.005	.027	-.15	.082
Lag(XVOL, -10)	217	.005	.027	-.15	.082
Lag(XVOL, -11)	216	.005	.027	-.15	.082
Lag(XVOL, -12)	215	.005	.027	-.15	.082
Lag(D(Unemployment), -1)	226	0	.001	-.002	.005
Lag(D(Unemployment), -2)	225	0	.001	-.002	.005
Lag(D(Unemployment), -3)	224	0	.001	-.002	.005
Lag(D(Unemployment), -4)	223	0	.001	-.002	.005
Lag(D(Unemployment), -5)	222	0	.001	-.002	.005
Lag(D(Unemployment), -6)	221	0	.001	-.002	.005
Lag(D(Unemployment), -7)	220	0	.001	-.002	.005
Lag(D(Unemployment), -8)	219	0	.001	-.002	.005
Lag(D(Unemployment), -9)	218	0	.001	-.002	.005
Lag(D(Unemployment), -10)	217	0	.001	-.002	.005
Lag(D(Unemployment), -11)	216	0	.001	-.002	.005
Lag(D(Unemployment), -12)	215	0	.001	-.002	.005
Lag(D(Spread Gov), -1)	226	0	.002	-.006	.01
Lag(D(Spread Gov), -2)	225	0	.002	-.006	.01
Lag(D(Spread Gov), -3)	224	0	.002	-.006	.01
Lag(D(Spread Gov), -4)	223	0	.002	-.006	.01
Lag(D(Spread Gov), -5)	222	0	.002	-.006	.01
Lag(D(Spread Gov), -6)	221	0	.002	-.006	.01
Lag(D(Spread Gov), -7)	220	0	.002	-.006	.01
Lag(D(Spread Gov), -8)	219	0	.002	-.006	.01
Lag(D(Spread Gov), -9)	218	0	.002	-.006	.01
Lag(D(Spread Gov), -10)	217	0	.002	-.006	.01
Lag(D(Spread Gov), -11)	216	0	.002	-.006	.01
Lag(D(Spread Gov), -12)	215	0	.002	-.006	.01
Lag(D(Spread Corp), -1)	226	0	.009	-.031	.071
Lag(D(Spread Corp), -2)	225	0	.009	-.031	.071
Lag(D(Spread Corp), -3)	224	0	.009	-.031	.071
Lag(D(Spread Corp), -4)	223	0	.009	-.031	.071
Lag(D(Spread Corp), -5)	222	0	.009	-.031	.071
Lag(D(Spread Corp), -6)	221	0	.009	-.031	.071
Lag(D(Spread Corp), -7)	220	0	.009	-.031	.071
Lag(D(Spread Corp), -8)	219	0	.009	-.031	.071
Lag(D(Spread Corp), -9)	218	0	.009	-.031	.071
Lag(D(Spread Corp), -10)	217	0	.009	-.031	.071

Lag(D(Spread Corp), -11)	216	0	.009	-.031	.071
Lag(D(Spread Corp), -12)	215	0	.009	-.031	.071

Table 10. *VIF multicollinearity test results of model regressors*

Model	Regressor	VIF	1/VIF
Model 2019	Lag(D(Euribor),-10)	1.012	0.988
Model 2019	Lag(IP, -5)	1.012	0.988
Model 2020	Lag(D(Euribor),-10)	1.038	0.963
Model 2020	Lag(D(Spread Corp), -8)	1.038	0.963
Model 2021	Lag(D(Euribor),-10)	1.035	0.966
Model 2021	Lag(D(Spread Corp), -8)	1.035	0.966
Model 2022	Lag(D(Euribor),-10)	1.032	0.969
Model 2022	Lag(D(Spread Corp), -8)	1.032	0.969
Model 2023	GFC (dummy)	1.534	0.652
Model 2023	Lag(D(Euribor),-10)	1.501	0.666
Model 2023	Lag(D(Spread Corp), -11)	1.259	0.794
Model 2023	Lag(D(Spread Corp), -8)	1.236	0.809
Model 2023	Lag(D(HICP), -11)	1.153	0.867
Model 2023	D(Bond Spread Gov)	1.130	0.885
Model 2023	Lag(D(Spread Gov), -1)	1.107	0.903
Model 2023	D(Corp Bond spread)	1.055	0.948

Table 11. *Tests on model residuals*

Model	Test Type	Statistic	Value	p-value
Model 2019	Autocorrelation (Portmanteau test, 5 lags)	Q statistic	6.8068	0.2354
Model 2019	Heteroskedasticity (Breusch-Pagan)	chi2(1)	0.67	0.4124
Model 2019	Normality (Shapiro-Wilk)	W	0.717	0.000***
Model 2020	Autocorrelation (Portmanteau test, 5 lags)	Q statistic	6.7531	0.1248
Model 2020	Heteroskedasticity (Breusch-Pagan)	chi2(1)	4.20	0.0405**
Model 2020	Normality (Shapiro-Wilk)	W	0.715	0.000***
Model 2021	Autocorrelation (Portmanteau test, 5 lags)	Q statistic	8.8222	0.1164
Model 2021	Heteroskedasticity (Breusch-Pagan)	chi2(1)	7.75	0.0054***
Model 2021	Normality (Shapiro-Wilk)	W	0.648	0.000***
Model 2022	Autocorrelation (Portmanteau test, 5 lags)	Q statistic	8.7455	0.1197
Model 2022	Heteroskedasticity (Breusch-Pagan)	chi2(1)	7.04	0.0080***
Model 2022	Normality (Shapiro-Wilk)	W	0.648	0.000***
Model 2023	Autocorrelation (Portmanteau test, 5 lags)	Q statistic	15.0537	0.0101***
Model 2023	Heteroskedasticity (Breusch-Pagan)	chi2(1)	39.05	0.0000***
Model 2023	Normality (Shapiro-Wilk)	W	0.764	0.000***

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Figure 2. *European Corporate Loan Default Rates over time*

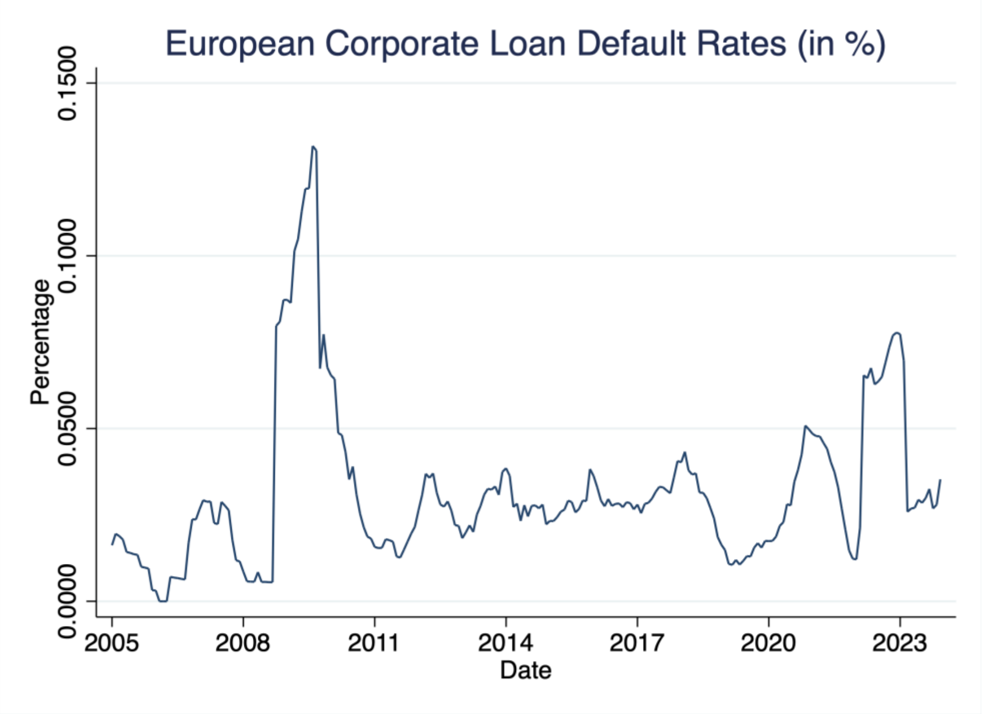

Figure 3. *3M Euribor interest rate over time (in %)*

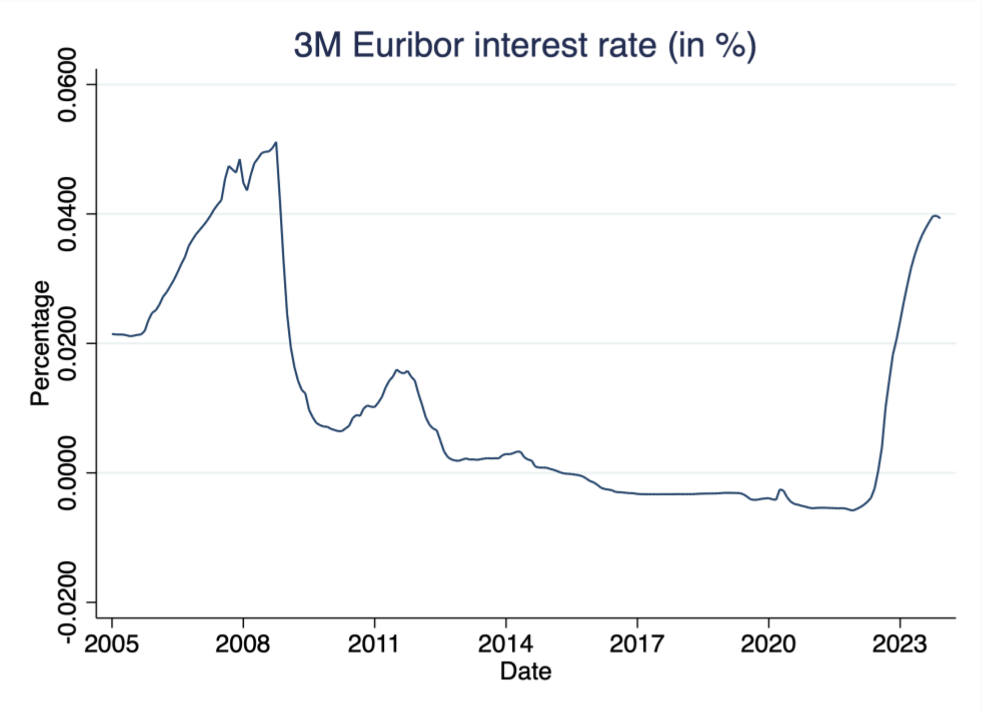

Figure 4. *European Industrial Production (in MoM % change)*

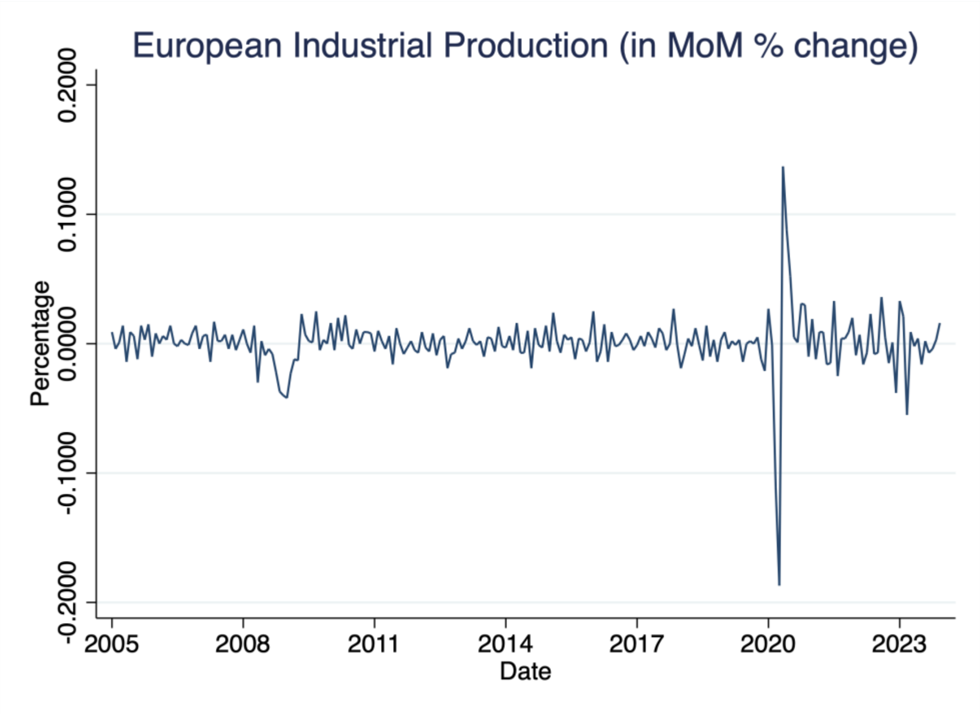

Figure 5. *European Harmonised Index of Consumer Prices (in %)*

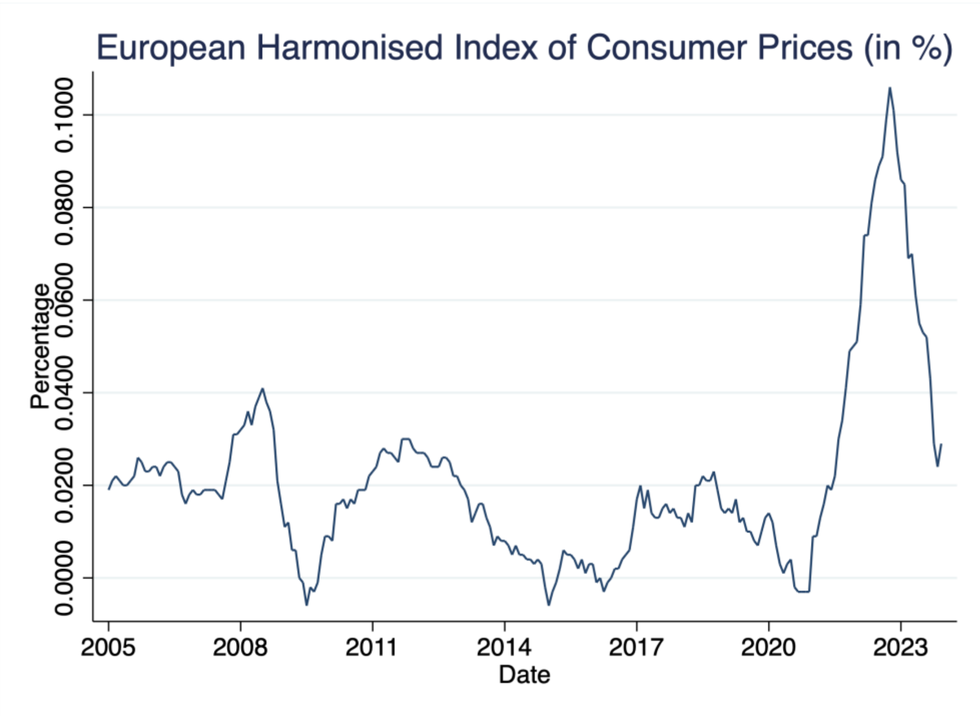

Figure 6. *European Traded Volume (in MoM % change)*

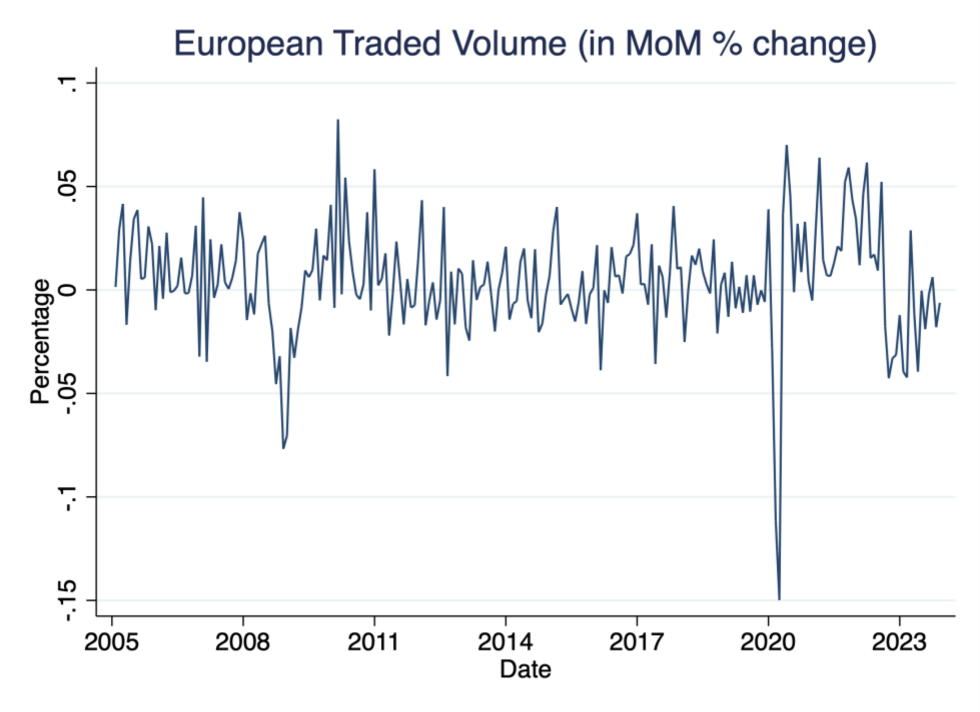

Figure 7. *European Unemployment Rate (in %)*

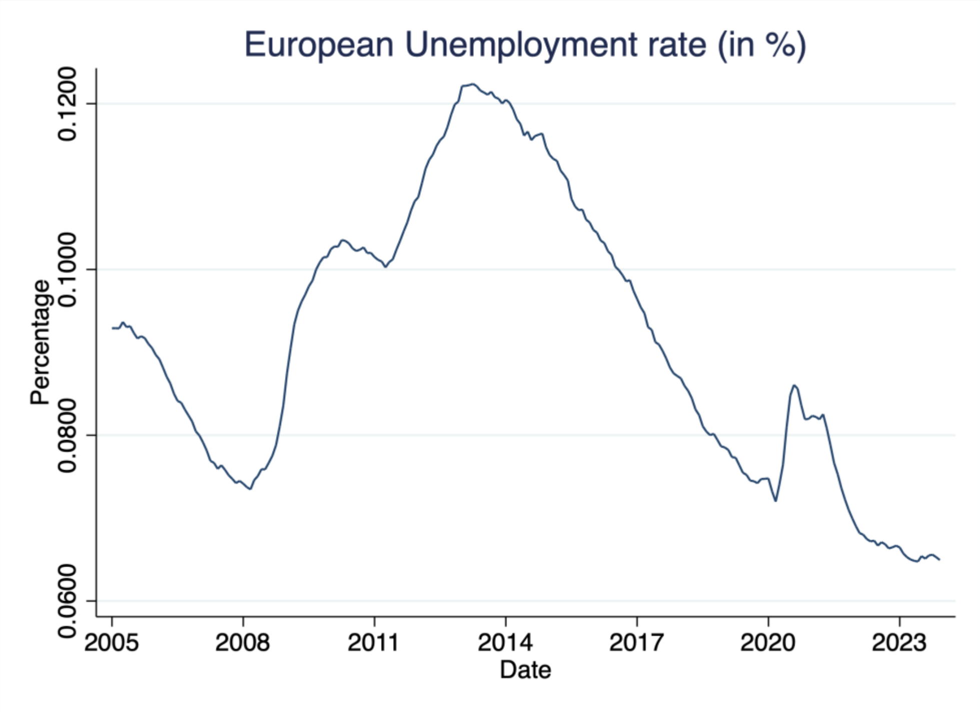

Figure 8. *European Governmental Bond Spread (in %)*

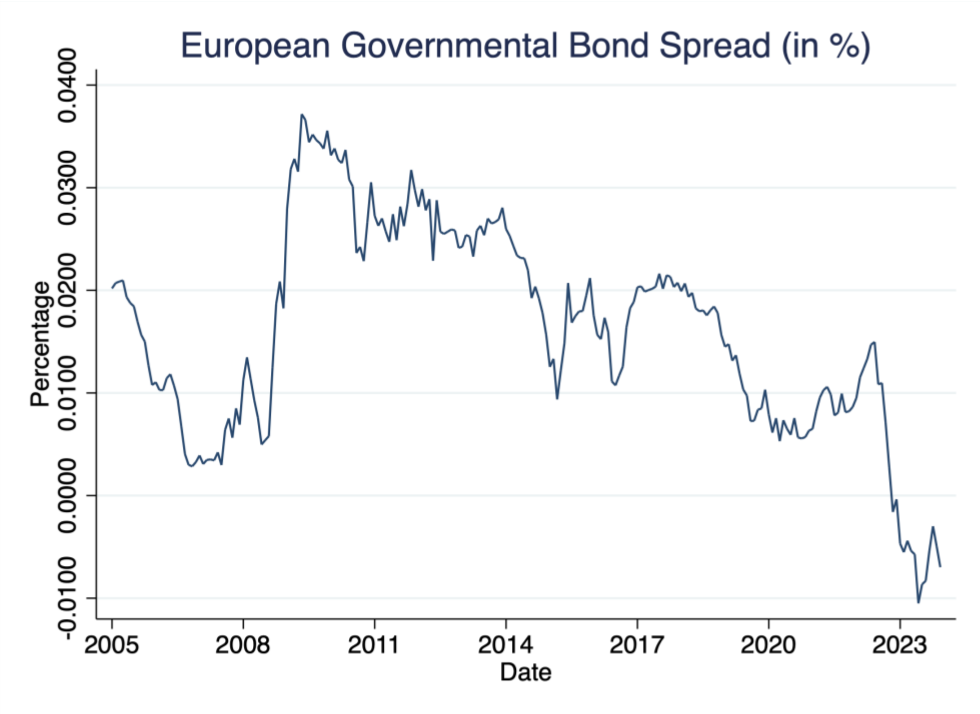

Figure 9. *European Corporate Bond Spread (in %)*