Master Thesis Quantitative Finance & Actuarial Science

A Survey of State-of-the-Economy Estimates for Point-In-Time Probability of Default Models

Author: Giel Sentjens (ANR: 650326, SNR: 2013808)

Supervisors: Bas J.M. Werker

Arwyn Goos (Deloitte), Jan Stekelenburg (Deloitte)

December 10, 2023

Tilburg University

Abstract: *This thesis evaluates various methods for estimating the 12-month Point-In-Time (PIT) Probability of Default (PD) using Freddie Mac mortgage data from 2000 to 2022. It compares the PIT PD estimates constructed using various estimation procedures for the "state-of-the-economy". The state-of-the-economy is estimated using three models: a Generalized Autoregressive Scoring (GAS) model, a Mixed Effects Logistic (MEL) model, and a model using a combination of observed macroeconomic characteristics. These PIT PD estimates are evaluated against a benchmark model, which assumes no influence from the state-of-the-economy, to identify the most effective method for estimating the PIT PD for U.S. mortgages. While including an estimate of the state-of-the-economy generally enhances model performance, the research is not able to determine a clear preference between the GAS model and the model using a combination of macroeconomic characteristics.*

Contents

1 Introduction

1.1 Preface

Following the aftermath of the 2008 financial crisis, the financial sector has intensified its focus on financial risk management. The primary factor commonly attributed to the onset of the 2008 financial crisis was the excessive amount of risk-taking by banks and other financial institutions. This can be exemplified by the fact that Collateralized Debt Obligations (CDOs) were often held within their investment portfolios. The omnipresence of CDOs is widely regarded as the primary cause of the financial crisis (see, for example, Crotty, 2009). These securities, essentially pooled mortgages, are complicated to such a degree that their valuation poses considerable challenges. During the mortgage crisis of 2008, banks incurred major losses and the presence of CDOs within their portfolios amplified this issue (Crotty, 2009). The disastrous consequences on households and financial institutions following at least in part from the difficulties in assessing the risks inherent to these CDOs highlighted the importance of credit risk monitoring and management in maintaining stability throughout the entire financial sector. Consequently, in the years after the crisis, a significant amount of new legislation was introduced regarding financial risk management. Most notably, Basel III was introduced to counteract the issues that led up to the financial crisis of 2008 (BCBS, 2010).

A central feature of the Basel framework is to determine the amount of risk capital a financial institution should maintain as a reserve. These reserves are crucial for mitigating solvency issues that could arise when an unexpected number of creditors fail to meet their repayment obligations. Alongside considering a range of economic factors to determine the necessary risk capital, the Basel guidelines also use detailed information on individual asset-level risks. A key quantity for this is the probability of default. This represents the probability that a borrower will default on their loan or debt, leading to financial losses for the lender. Within the literature on the modeling of the probability of default, two different classes of models are commonly employed. These are categorized as *Through-The-Cycle* (TTC) and *Point-In-Time* (PIT) models. Both have different and complementary roles from a credit risk perspective.

TTC estimates are designed so that the average Observed Default Rate (ODR) over longer periods matches the TTC default rate prediction. In particular, a TTC estimate of the PD is designed such that it is only slightly affected by the underlying fluctuations of the business cycle or default cycle. These underlying fluctuations will be referred to as the “state-of-the-economy” throughout this thesis. Because TTC estimates are designed in this way, they vary little over time. A characteristic of this approach is that it prevents risk capital from fluctuating substantially, which might be desirable in some situations. Based on this rationale, the regulation following from Basel accords starting from Basel II (BCBS, 2006) requires default probabilities to be TTC.

Due to their design, TTC models for estimating the probability of default lack flexibility in the sense that the estimated default rate may not reflect the realized default rate in a shorter horizon. When more loans default than would be expected from the TTC PD estimate, the level of risk capital may not be sufficient and the bank may encounter solvency issues. On the other hand, in years where economic conditions are favorable, the amount of risk capital may be unnecessarily large preventing it from being used elsewhere in the bank. In particular, the potential downside of TTC estimates of the PD became apparent during the financial crisis, leading to PIT models for estimating the probability of default to be more widely adopted and required by regulatory frameworks.

PIT estimates for the probability of default are constructed differently than TTC estimates. Rather than being invariant to the state-of-the-economy, PIT estimates are constructed to reflect this state-of-the-economy, implicitly assuming that the state-of-the-economy and the default rate are correlated. While this assumption is subject to debate, the seminal textbook on credit risk by Duffie & Singleton (2012) affirms it. Further validation can be found in for example Koopman & Lucas (2005a) and Castro (2013).

Effectively, the inclusion of the state-of-the-economy results in PIT PD estimates reflecting the ODR more closely. Due to this inclusion, PIT PD estimates exhibit larger fluctuations over time than their TTC counterparts, which is thereby also reflected in the required amounts of provisions. When estimated correctly, these provisions represent the true requirements more accurately, but managing the associated fluctuations requires additional effort and resources. In addition to the complexity of managing these fluctuations, the estimation of PIT PDs itself is not straightforward. PIT PDs can be estimated in a variety of ways while still complying with regulatory requirements. Because of this wide range of possible models and the relevant regulation being relatively recent, it is unclear in what configuration a PIT PD is most reliably estimated. In Figure 1 the difference between the two concepts is illustrated.

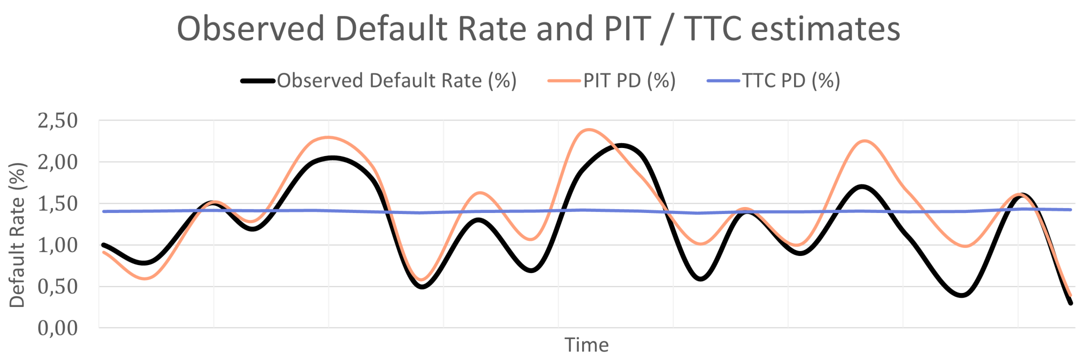


Figure 1: *An example figure highlighting the difference between TTC and PIT estimates of the PD for a portfolio of loans. The black line represents the Observed Default Rate, the yellow line is a PIT PD estimate of the default rate and the blue line is a TTC PD estimate. The key take-away is that over time the TTC estimate is notably less variant than the PIT estimate.*

Briefly put, by taking into account the current state-of-the-economy, PIT PD estimates offer a more dependable estimate of the actual default rates at that specific moment, but this increased preciseness also increases variability. PIT PD models are required by the latest international financial reporting standard, IFRS9, and EBA regulations on stress testing procedures (EBA, 2022a) and often use macroeconomic factors like unemployment or GDP growth rates. This thesis specifically focuses on the application of PIT PD models to mortgage default probabilities. Mortgages, due to their significant share in the portfolios of financial institutions, are a natural case study for examining the effectiveness and limitations of PIT PD models in varying configurations.

While PIT PD models predominantly utilize macroeconomic factors to reflect the state-of-the-economy, these macroeconomic factors may not always yield accurate estimates for the PD of mortgages. This was particularly evident during the COVID-19 crisis in the EU. Despite adverse macroeconomic indicators, default rates did not increase as strongly as would be expected, presumably due to the extensive government subsidies provided during this period (see for example Dutch Ministry of Finance (2022), EBA (2020)). That these subsidies are effective is substantiated by evidence from the UK, as detailed in Albuquerque & Varadi (2023), and the broader effectiveness of financial relief measures during the COVID-19 crisis is confirmed in Biljanovska et al. (2023). Based on this, it is suggested that also political factors might affect mortgage default rates, in addition to mortgage-specific or macroeconomic characteristics.

This observation highlights a key limitation of traditional PIT PD models: their reliance on macroeconomic variables may thus not accurately mirror the true PD. To address this shortcoming, two alternative approaches to estimate the PIT PD will be explored in this thesis. These are both econometric techniques that provide a different modeling of the state-of-the-economy, aiming to more accurately capture the complexities of PD estimation under unforeseen fluctuations, such as during the COVID-19 crisis. These broader approaches might offer a more robust way for PD modeling, especially in scenarios where traditional macroeconomic indicators do not fully capture the underlying financial dynamics.

1.2 Research Goals

This thesis aims to survey and assess the performance of different PIT models, each estimating the 12-month probability of default for individual mortgages and estimating the state-of-the-economy in different ways. The primary goal of this study is to identify the model that most accurately estimates the PIT PD, thereby minimizing the mismatch between the estimated required amount of provisions and the actual amount that is required. The analysis will be conducted on a portfolio of US single-family mortgages, all of which were active for a certain period between the years 2000 to 2022. The data on these mortgages is obtained from Freddie Mac, a US government-sponsored enterprise that buys mortgages, pools them, and then sells these pooled mortgages as mortgage-backed securities to investors.

Using these mortgage data, the PIT PD estimates will be determined for the various methodologies. The subsequent analysis will then be to assess the estimates obtained from the PIT models by comparing the estimated default probabilities against default realizations.

Four models are included in this survey, all of which rely on logistic regression to estimate the probability that a creditor will default within a time frame of 12 months, hereafter simply referred to as the 12-month probability of default. The details for all these models are outlined in Section 3. Additionally, an intercept-only model will be included for further reference of the models' performance.

The first model exclusively uses observable mortgage-specific risk drivers for the probability of default estimate and serves as the benchmark model.

In the second model, macroeconomic variables such as the unemployment rate and the GDP growth rate will be included. These macroeconomic factors are all retrieved from the US Bureau of Economic Analysis[1].

In the third model, in each period the state-of-the-economy is randomly sampled from a normal distribution. After some manipulation, the period-specific states of the economy can nevertheless be derived so that one can estimate whether the number of defaults in that period should either be relatively high or low. An advantage over the other two models is that while this model does not need to observe macroeconomic characteristics directly, it is able to reflect the macroeconomic conditions of a given year. This is modeled using a Mixed Effect Logistic model, which is thoroughly discussed in Demidenko (2013) and Mcculloch & Neuhaus (2013).

In the fourth model, the state-of-the-economy is modeled using a *Generalized Autoregressive Score* (hereafter, *GAS*) process, originally described in Creal et al. (2013). In this model, the state-of-the-economy follows a recursive relation. The updating steps of this recursion depend on the data observed in each given period. One advantage of this model lies in reflecting the macroeconomic conditions of a given year without the need for direct observation of macroeconomic characteristics. It also imposes a form on the state-of-the-economy that is arguably less restrictive than the MEL model.

The four models shall then all be evaluated by fitting and testing these on the Freddie Mac data in order to determine which of the four models is best suited for estimating PIT PDs for US mortgages. For all four models, the estimated default probabilities shall be compared to the realized defaults. The main target of evaluating these approaches shall be to find out which model is best suited to provide a PIT PD estimate. In this comparison, model performance in terms of calibration accuracy shall be assessed by means of the mean squared error (MSE). In order to assess the discriminatory power, the Area Under the Curve (AUC) shall be computed. These metrics shall then be used to compare the performances of the PIT models. Moreover, various likelihood ratio tests shall assess whether the effect of the state-of-the-economy on the default rate is significantly different from zero or not.

The objective of the performance analysis is to investigate the potential of the modeling methodologies for estimating the PIT PD. The main question this thesis shall aim to address is therefore:

Research Question*: Which of the following methods for estimating the state-of-the-economy yields the most reliable Point-In-Time PD model for U.S. mortgages: using a linear combination of macroeconomic indicators, implementing a GAS recursion, using a Mixed Effects Logistic model, or simplifying the model by not incorporating any state-of-the-economy estimate?*

In order to answer this research question, the remainder of this thesis will be structured as follows: In Section 2 a theoretical background is provided, highlighting some background information on credit risk and introducing more precise definitions of PIT and TTC PDs. In Section 3, the methodologies used the compute the default probability will be discussed. In Section 4, the estimation procedures for the models will be discussed. Moreover, the performance measurement techniques used to answer which model is

[1]The macroeconomic data by the US BEA are retrieved from the website of the FRED economic data: https://fred.stlouisfed.org/

"most reliable" will be outlined. Section 5 will discuss the Freddie Mac and the macroeconomic data, including summary statistics and details on data collection and cleaning processes. In Section 6 the results from the estimation procedures will be presented, together with tables and figures summarizing these. Lastly, in Section 7 the results will be discussed and some comments regarding the potential implications of the various PIT models are provided. Furthermore, some research limitations and openings for future research will be provided in this section.

2 Theoretical Context

2.1 Credit Risk Management: Overview and Contextual Framework

Before proceeding with the more detailed theoretical discussion in this thesis, some general context regarding the importance of both financial risk management and credit risk management and how these concepts interconnect is provided. Although the PIT probability of default models are primarily relevant from an IFRS9 and stress-testing perspective, the discussion of financial risk management below will mostly be based on the guidelines by the Basel Committee on Banking Supervision (BCBS). Adopting this perspective provides a clearer understanding of the importance of credit risk management and emphasizes the significance of a systematic estimation of the probability of default, given that the BCBS serves as the overarching regulatory framework to which, be it indirectly, IFRS9 and EBA regulation are subject.

In financial risk management, five types of risk are commonly identified, as extensively discussed in Roncalli (2020). Market risk is defined as the risk of losses due to fluctuations in market price. Counter-party credit risk is another category, relating to the possibility of default in derivative instruments like swaps or options, and is thereby distinct from general credit risk. The third risk category is operational risk, encompassing the potential for losses arising from insufficient or unsuccessful internal processes, personnel and systems, as well as external incidents. Examples of operational risk are fraud or natural disasters. The fourth type, liquidity risk is the risk of losses resulting from the failure of the financial institution to meet its payment obligations on time. Finally, credit risk is the risk of loss arising from a counterparty's non-fulfillment of financial obligations. This risk will be central throughout this thesis and will be described in further detail later in this section.

In order to effectively manage these risks, Basel guidelines provide extensive legislation. The Basel Committee on Banking Supervision (BCBS), founded in 1974, plays a central role in developing and implementing these regulations. The first accord, Basel I, was issued in 1988. During these early years, the BCBS stated its objective as enhancing financial stability through the advancement of supervisory expertise and the overall quality of banking supervision around the world. Presently, the focus has drifted somewhat from its original aim, which is currently more aimed at monitoring and ensuring capital adequacy (BCBS, 2023). The Basel guidelines have undergone multiple revisions over the course of the years and the most recent issue which is currently in effect is Basel III[2]. Therefore, the discussion below shall be based on Basel III.

Basel III distinguishes between three pillars which collectively aim to ensure the capital adequacy and stability of banks. The first of these pillars is Pillar I, which provides requirements for minimum capital and liquidity. This pillar is the most relevant for this thesis. Pillar II describes a supervisory review process under which supervisors review banks' own assessments of their capital adequacy, thereby providing a rigorous double-check if the bank is meeting its capital requirements. Pillar III requires banks to be transparent by publicly disclosing capital positions and their market, credit and operational risk exposures. Most of the contents discussed in this thesis are almost exclusively relevant from a Pillar I perspective. Therefore, Pillar II and III are out of scope for this thesis and only Pillar I will be discussed in further detail.

In order to provide capital requirements based on Pillar I, Basel III provides an extensive methodological outline for all five types of financial risk. Despite the restrictions on the procedures to determine the amounts of risk capital, a bank does possess a degree of latitude for using internal models and having an influence on the amount of risk capital. The size of this capital buffer is largely determined by the amount of the risk-weighted asset exposures. These are the asset exposures held by the bank, each weighted by the risk weighting factor for that asset class. For example, the weighting factor for cash

[2] It should be mentioned that as of January 1st, 2023 transitioning to Basel IV has commenced.

is lower than the weighting factor for stocks. To illustrate this further in an extreme and stylized scenario, a bank that keeps its entire asset portfolio in cash would require significantly less capital to offset risks compared to a bank that holds all its assets in stocks. From the Basel regulation, it is required that at least a certain percentage of the risk-weighted value of these assets should be used as capital buffer.

Under the Basel III guidelines, banks may calculate these risk-weighted assets either via the Standardised approach or the internal ratings-based (IRB) approach. In the Standardised approach, the risk weights are fixed per asset category. The advantage of this is that it saves the bank time and resources in determining this quantity. One potential drawback of the Standardised approach is that it may lead to larger estimates of risk-weighted assets compared to what would have been calculated if the unique characteristics of the bank were taken into account. It is possible that this approach might result in the bank maintaining a larger capital buffer than is strictly required for meeting the Basel guidelines.

While IRB approaches are customized to a bank's specific needs, their development and implementation involve more effort, mainly because of the extensive legislative guidelines the bank must follow. In order to determine the capital requirement, banks use the risk-weighted asset exposure mentioned earlier. This formula varies per asset category. Below, the formula used by the EBA for consumer mortgages (EBA, 2022b) is included.

$$RW = \left(LGD \cdot \Phi\left(\frac{1}{\sqrt{1-R}} \cdot \Phi^{-1}(PD) + \sqrt{\frac{R}{1-R}} \cdot \Phi^{-1}(0.999) \right) - LGD \cdot PD \right) \cdot 12.5 \cdot 1.06. \tag{1}$$

Here, RW is the risk weight of the asset, $\Phi(\cdot)$ is the standard normal cumulative distribution function, LGD is the loss given default of that particular asset and PD its probability of default. Both LGD and PD are determined using specific modeling approaches. R is the "coefficient of correlation". For brevity, this definition is deferred to Equation (59) in Appendix B.1.

In addition to the guidelines on internal solvency requirements outlined by the Basel Committee, IFRS9 serves as a framework for financial accounting and not specifically for banking in general. Unlike the IRB approach of Basel III, which calculates risk-weighted assets, IFRS9 determines provisions based on the Expected Loss (EL), a different metric. The Expected Loss is defined as

$$EL = PD \times LGD \times EAD. \tag{2}$$

So, the Expected Loss is the product of the Probability of Default (PD), the Loss Given Default (LGD) and the Exposure At Default (EAD). All three of these quantities should be modeled separately. Moreover, the way in which the PD and LGD are calculated also differs from those of the IRB approach.

This thesis primarily focuses on modeling the PD, appearing in both Equation (1) and Equation (2). As previously indicated, these equations have a similar but not the same application: IRB models use the risk-weighted asset formula in Equation (1) to determine the amount of risk capital and require PDs to be TTC. Conversely, the IFRS9 framework bases its provisions on the Expected Loss in Equation (2), where PD estimates are typically PIT. Therefore, it can be concluded that PIT and TTC PDs serve similar but not equivalent purposes. To understand the relation of these two philosophies, in the next subsection, a general definition of the PD shall be provided. This will set the basis for a detailed comparison of the TTC and PIT approaches to PD modeling.

2.2 Point-in-Time and Through-the-Cycle PD Definitions

2.2.1 General PD Definition

In order to provide an outline for the probability of default estimation, definitions for the PIT probability of default as well as for the TTC probability of default will be introduced in this section. The TTC definition shall not be used in the analysis later, but rather serves to provide some more clarity on the difference between the two philosophies. Before proceeding with these definitions, first a general definition of the PD will be provided here.

Suppose a bank holds a portfolio of $j = 1, ..., J$ loans and these are observed in some or all periods in $t = 1, ..., T$. The objective is to model the probability of default for each individual loan. To start, define for all loans j in all periods t a default indicator D_t^j. To model the probability of default effectively, assume for now that the underlying distribution function of D_t^j is dependent on loan-specific risk drivers, denoted by $X_t^j = \{X_{1t}^j, X_{2t}^j, \ldots, X_{kt}^j\}$, as well as on broader economic conditions described by u_t. For notational simplicity, Z_t^j is defined as set the of predictors that combines these risk drivers with the state-of-the-economy such that: $Z_t^j = \{X_{1t}^j, X_{2t}^j, \ldots, X_{kt}^j, u_t\}$.

In probability of default modeling, the typical quantity of interest is the probability that a loan j is in default h periods after period t. Since only default realizations up until period t are observed, this requires estimation. To now define the probability of default, write

$$p_{t,t+h}^j = \mathbb{P}(D_{t+h}^j = 1 \mid Z_t^j). \tag{3}$$

Here $p_{t,t+h}^j$ is the probability that the loan defaults within h periods Z_t^j, conditional on the observed characteristics of loan j in period t. As highlighted in the previous Subsection in Equation (1) and Equation (2), this quantity is crucial for IRB and IFRS9 modeling and thereby also for credit risk modeling in general.

2.2.2 Point-In-Time PD Definition

As already noted, there exist two philosophies with different views on how the PD should be estimated, being the Through-the-Cycle and the Point-in-Time models. This dualism has been the center of debate for decades. Early definitions of these philosophies are in Treacy & Carey (1998), while Eder (2021) offers a comprehensive overview of the various definitions proposed for both. In this thesis, the starting point to define these notions will be that the PIT definition of the PD reflects the current state-of-the-economy, whereas the TTC PD is robust against this state-of-the-economy. The term “state-of-the-economy” does not exclusively refer to how the economy affects the default rate. Instead, it encompasses the total influence of systematic variations in the default rate that are not captured by the risk drivers in the models. This includes, but is not limited to, various macroeconomic factors. For example, political or environmental factors might also impact the default rate.

In Chawla et al. (2017), the regulatory context of PD estimation is carefully described. The information provided in this subsection shall be based on that paper. As mentioned in Section 2.1, PIT PDs are used in both IFRS9 reporting and stress-testing. In paragraph 85 of EBA’s stress-testing-methodology document (EBA, 2022a), this PIT requirement is mentioned. There, it is stated that: “in all credit-risk-related calculations except RWA for all portfolios, institutions should use point-in-time (PIT) measures that reflect the current outlook for business-cycle conditions”.

Furthermore, in the IFRS9 requirements for financial instruments (IASB, 2021) this requirement is also clearly mentioned. There, paragraph 5.5.17 states: “An entity shall measure expected credit losses of a financial instrument in a way that reflects an unbiased and probability-weighted amount that is determined by evaluating a range of possible outcomes and reasonable and supportable information that is available without undue cost or effort at the reporting date about past events, current conditions and forecasts of future economic conditions.” Of course, in particular, the latter remark is reminiscent of the reason why the PIT PD is examined in this thesis.

Having considered the regulatory requirements of both the EBA and IFRS9, the relevance of PIT PDs is clear. To capture cyclical or macroeconomic effects required for PIT PDs in these frameworks, a proxy term for the state-of-the-economy u_t will be introduced in each of the models that will be evaluated. The interpretation of this quantity u_t is that in each period, the Observed Default Rate might be subject to some unobserved fluctuations. These fluctuations can be but are not necessarily restricted to economic and political fluctuations. The effect of such a fluctuation on the observed default may be challenging to describe by means of a linear combination of macroeconomic characteristics. In Section 3, various definitions of u_t shall be provided.

To formalize how this “state-of-the-economy” interacts with the Observed Default Rate, the PIT PD in this thesis will be defined as a probability of default conditional on the state-of-the-economy. It should be stressed once more that there is no universally accepted definition of the PIT PD and this also applies to the definition presented here. The definition is selected for its simplicity and potential to provide insights within various PD modeling methods, while still aligning with the descriptions of the concept provided by regulatory frameworks. The same definition of the PIT PD is also used in for example Miu et al. (2005), Tasche (2006) and Carlehed & Petrov (2012). Write for the definition of the PIT PD

Definition 1a) The Point-In-Time Probability of Default is the probability that a loan j will default between time t and $t+h$, conditional on the set of predictors $Z_t^j = \{X_{1t}^j, X_{2t}^j, \ldots, X_{kt}^j, u_t\}$. Here, X_t^j is the set of observed loan-specific characteristics and u_t the state-of-the-economy. The Point-In-Time Probability of Default can then be written as

$$p_{t,t+h}^{j,PIT} = \mathbb{P}(D_{t+h}^j = 1 \mid Z_t^j). \tag{4}$$

2.2.3 Through-The-Cycle PD Definition

Unlike PIT PDs, TTC PDs are characterized by their stability throughout economic cycles. Although in this thesis the TTC PD will not be applied directly to assess model performance, the following discussion aims to provide some more clarity on the PIT-TTC dualism.

As previously discussed, TTC approaches to estimating the PD are required for IRB modeling. There, PDs are required to be TTC when computing risk-weighted assets for the IRB approach, as shown in Equation (1). Within the Basel framework itself, this is specified in the legal guidelines detailed in chapter 36 for credit risk (BCBS, 2019), labeled CRE36[3].

In section 77 of that chapter (CRE 36.77), it is noted that banks are required to use techniques that consider long-term experience when estimating the PD. By CRE36.78 banks must estimate a PD based on the observed historical average of the one-year default rate. This average does not fluctuate as strongly over time as an individual 12 month-default rate would, amounting to a more stable estimate of the PD.

Considering these regulatory CRE requirements, it can be said that from an IRB perspective PDs are required to be stable over time and reflect a long-term average PD or at least a PD mostly unaffected by the state-of-the-economy of that given year.

More specifically, within this thesis, the TTC PD is defined as the probability of default unconditional to state-of-the-economy u_t. Again, this is not a generally accepted definition, but it will be maintained for the purposes of this study. By using this definition, the TTC PD averages out the PD over macroeconomic and structural fluctuations, thereby aligning well with the Basel III guidelines for robustness to short-term economic fluctuations. Define the TTC PD as

Definition 1b) The Through-The-Cycle probability of default is the probability that a loan j will default between time t and $t+h$, conditional on the set of predictors $Z_t^j = \{X_{1t}^j, X_{2t}^j, \ldots, X_{kt}^j\}$. Here, X_t^j is the set of observed loan-specific characteristics and the state-of-the-economy u_t is not in Z_t^j. The Through-The-Cycle Probability of Default can then be written as

$$p_{t,t+h}^{j,TTC} = \mathbb{P}(D_{t+h}^j = 1 \mid Z_t^j) \text{ and } u_t \notin Z_t^j \tag{5}$$

In Section 3, Definition 1a shall be used in order to find how to estimate the 12-month probability of default as reliably as possible among the specifications that will be employed here. The starting point will be a PIT model which estimates the 12-month probability of default exclusively based on loan characteristics. Then, when this model has been calibrated, this shall be shifted based on the u_t introduced above.

Before finishing the discussion on the PD definitions, what should be mentioned is that PIT PD models are often developed from existing IRB models, which provide TTC estimates. In practice, the degree of "PIT-ness" in PD models often varies. When an IRB TTC PD model has been developed, banks determine the extent to which they adjust the TTC PD model to reflect economic conditions when constructing PIT PDs, effectively controlling its "PIT" intensity.

[3] Calculation of Risk-Weighted Assets Chapter 36

3 Methodology

To estimate the 12-month probability of default, various approaches will be employed in this thesis. These approaches will be introduced in this section, which will be structured as follows: In Subsection 3.1 the general setting which is used to estimate the default probabilities is described. In Subsection 3.2.1 the benchmark model is discussed. In this benchmark model, it is assumed that there are no structural fluctuations of the default rate, so u_t is assumed to be zero in all periods. The models discussed in the subsections thereafter shall expand upon the benchmark model by including a structural error term u_t. In Subsection 3.2.2 u_t shall be a linear combination of macroeconomic data. In Subsection 3.2.3 u_t shall assumed to be an *i.i.d.* normally distributed random variable in each period t, but after some manipulation, u_t can still be estimated. In Subsection 3.2.4, u_t shall be defined by a GAS recursion. All models are estimated using maximum likelihood in *PYTHON*. The details of the estimation process and corresponding likelihood expressions are provided in Section 4. This section will primarily concentrate on defining the models.

3.1 General Setting

In this subsection, the general setting under which the mortgage default probabilities are modeled is outlined. The assumptions provided at the end of this subsection will serve as the framework for the rest of the models applied in this thesis.

First consider the structure of the data that is considered. There are T periods $t = 1, 2, ..., T$ in which J mortgages $j = 1, ..., J$ are observed. In each period t, a set of active mortgages L_t is observed, containing some subset of the mortgage indices $j = 1, ..., J$. The observations from the set of active mortgages $L_{t'}$ from period $t' < t$ are also known at each subsequent period t. Some mortgages in set L_t might also be in the set for the next period L_{t+1}, while others might leave the set because they are paid off or are in default. In other words, the sets mortgages from past periods are known at time t and this set of sets is denoted $\mathcal{L}_t = \{L_1, \ldots, L_t\}$ for periods $t = 1, \ldots, T$. A stylized representation of this structure is included in Figure 2 below. In each period t new mortgages may enter the set. Each of these mortgages has an observed default indicator D_t^j and loan characteristics X_t^j. Additionally, there is a systematic error term u_t affecting all mortgages in each period. This error term is independent of all loan characteristics X_t^j.

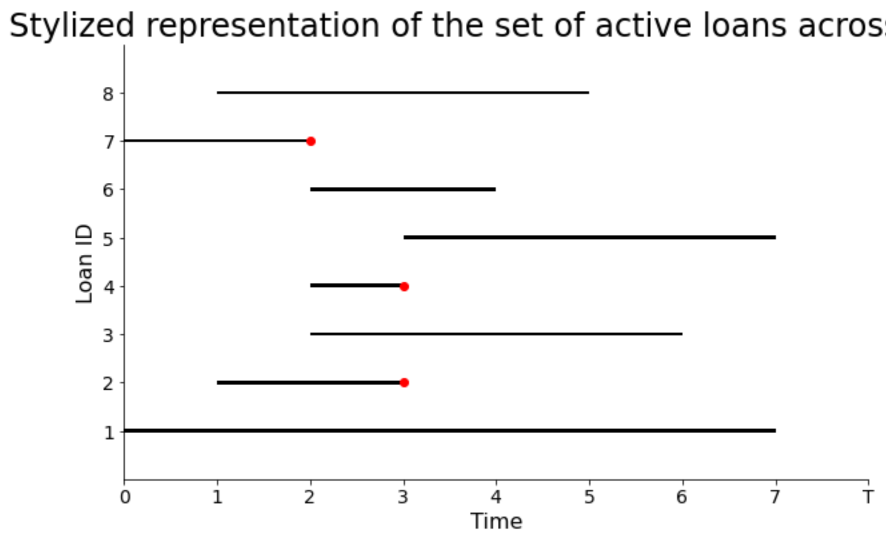


Figure 2: *Stylized representation of the structure of the mortgage data. The horizontal lines indicate the life span of an individual mortgage. The red dots mark a default event of a mortgage.*

The default indicator D_t^j equals 1 if mortgage j is in default in period t and 0 otherwise. However, the quantity of interest is not whether mortgage j is in default now, but rather if mortgage j will go into default between period t and period $t + h$. Throughout this thesis, h is set equal to 12. To model this, assume that default indicator $D_{t,t+12}^j$ equals 1 if mortgage j is in default between period t and $t + 12$ and that this indicator is Bernoulli distributed as in

$$D^j_{t,t+12} \sim Bernoulli(p^j_{t,t+12}). \tag{6}$$

Before defining the probability of default, one auxiliary assumption shall be made. This is the assumption that loans defaulting between t and $t+h$ can then be defined by assuming that loans can no longer recover once in default. In other words,

$$\mathbb{P}(D^j_{t+h} = 1 \mid D^j_t = 1) = 1. \tag{7}$$

While this assumption may only be partially true, there is a case to be made by noting that loans that have defaulted before might have statistically different characteristics than those that have not. Therefore, this assumption will be maintained throughout this thesis. This is verified in Table 7 in Appendix C.1.

By this auxiliary assumption, it is not relevant to include mortgages already in default at time t, since by this definition they must also be in default in period $t+h$. Consequently, the notation in Equation (6) can be simplified from $D^j_{t,t+12}$ to D^j_{t+12}. Next, for the underlying model of the observed default indicator D^j_{t+12} write

$$D^j_{t+12} = I\left(D^{j*}_{t+12} > 0\right) = I\left(\beta_0 + \sum_{i=1}^{k} \beta_i X^j_{i,t} + u_t + \varepsilon_{jt} > 0\right). \tag{8}$$

Throughout this thesis, for all models, it shall be assumed that ε_{jt} from Equation (8) is i.i.d. standard logistically distributed. The standard logistic distribution is given by

$$\Lambda(x; 0, 1) = \frac{1}{1 + e^{-x}}. \tag{9}$$

Therefore, the conditional probability that mortgage j will default at some time between t and $t+12$ is given by $p^j_{t,t+12} = \mathbb{P}(D^j_{t+12} = 1 \mid u_t, X^j_t, D^j_t = 0)$ and can be written as

$$p^j_{t,t+12} = \mathbb{P}(D^j_{t+12} = 1 \mid u_t, X^j_t, D^j_t = 0) = E(D^j_{t+12} \mid u_t, X^j_t, D^j_t = 0). \tag{10}$$

Under this logistic specification of the error term, Equation (8) and Equation (10) can be combined to find

$$logit(p^j_{t,t+12}) = Log\left(\frac{p^j_{t,t+12}}{1 - p^j_{t,t+12}}\right) = \beta_0 + \sum_{i=1}^{k} \beta_i X^j_{i,t} + u_t. \tag{11}$$

3.2 Model Specifications

3.2.1 Benchmark PD Model

In this subsection the benchmark PIT model shall be described. This will serve as the starting point of the other PIT models which include u_t. In this model, it shall be assumed that u_t from Equation (11) equals zero across all periods.

$$u_t = 0, \text{ for all } t = 1, ..., T. \tag{12}$$

The economic interpretation of $u_t = 0$ is that across all periods there are no structural fluctuations on the probability of default can be estimated with loan characteristics only. Excluding this error term allows for the possibility of assessing the value gained by including it. In this specification, one can write

$$logit(p^j_{t,t+12}) = Log\left(\frac{p^j_{t,t+12}}{1 - p^j_{t,t+12}}\right) = \beta_0 + \sum_{i=1}^{k} \beta_i X^j_{i,t}. \tag{13}$$

3.2.2 PD Model with Explicit Macroeconomic Characteristics

A straightforward way with which to include a "systematic" error term affecting all default probabilities might be by attempting to construct a value that serves as a proxy of the state-of-the-economy at the beginning of that period. These can then enter the model equation as additional risk drivers. Examples of these are the unemployment rate and the GDP growth rate. The macroeconomic risk drivers that will be used in the final analysis are discussed in more detail in Section 5.3. Each macroeconomic variable is denoted by M_{it} with $i = 1, ..., k_m$ its respective index. Write

$$logit(p^j_{t,t+12}) = Log\left(\frac{p^j_{t,t+12}}{1 - p^j_{t,t+12}}\right) = \beta_0 + \sum_{i=1}^{k} \beta_i X^j_{i,t} + \sum_{i=1}^{k_M} \gamma_i M_{it}. \tag{14}$$

3.2.3 Mixed Effects Logistic PD Model

While risk drivers are often sufficient for modeling default probabilities, they may not accurately describe all fluctuations in default rates. This is thoroughly described in the paper by Koopman et al. (2011), where more than 100 macroeconomic variables are used to model defaults but a significant part of the variance in default rates remains unexplained. As mentioned in the introduction, the limited increase in the PD of mortgages of households during COVID-19 is a good example of this. So, even after collecting and selecting macroeconomic risk drivers, there is likely still room for improvement in modeling the PD.

To accommodate the possibility of a systematic error term affecting all mortgages more flexibly, assume that $u_t \sim N(0, \sigma^2)$ for all $t = 1, ..., T$. By assuming that u_t follows a normal distribution, this specification acknowledges the presence of unobserved factors that may impact loan defaults but are not explicitly captured by the selected risk drivers. In addition to the advantage that a normally distributed state-of-the-economy yields in the sense that macroeconomic risk drivers are no longer required, one can assess whether there are any structural effects present by testing whether $\sigma = 0$ by comparing the benchmark model to this model. This PD model is no longer simply a logistic model, because additional to the error term ε_{jt} from Equation (8) there is now a second error term. Therefore, it should be referred to as a Mixed Effects Logistic (hereafter, MEL) model.

A MEL Model is a specific type of Generalized Linear Mixed Model (GLMM), in the same way that a logistic model is a specific type of Generalized Linear Model (GLM). In such a model, an additional random effect may be added to a GLM model equation as is done in Equation (15) below. For an extensive discussion regarding the properties of GLMM models, one can refer to for example Mcculloch & Neuhaus (2013) or Demidenko (2013).

Using the definitions in these books, the difference between the two model categories can be summarized as follows. In a GLM, the objective is to model the expected value of Y given X, represented as $\mathrm{E}(Y \mid X) = g^{-1}(X\beta)$, where g denotes the link function. Throughout this thesis, the link function is the logistic distribution function, as provided in Equation (9). On the other hand, a GLMM extends this framework by writing $g(E[y \mid X, u]) = X\beta + Zu$. This addition accommodates the inclusion of random effects or unobserved variables in the model, represented by u. Although in principle it is possible to add a vector of random effects u, in this thesis u_t is assumed to be one-dimensional. This effect is then added to the PD definition, so that the MEL PD model equation becomes

$$logit(p^j_{t,t+12}) = Log\left(\frac{p^j_{t,t+12}}{1 - p^j_{t,t+12}}\right) = \beta_0 + \sum_{i=1}^{k} \beta_i X^j_{i,t} + u_t \ \ with\ u_t \sim N(0, \sigma^2)\ \ i.i.d.\ and\ unobserved. \tag{15}$$

Part of the reason for assuming a normally distributed u_t relates to the specific estimation process of the MEL model. This includes using a Gauss-Hermite numerical approximation in the likelihood objective function. The fact that u_t is normally distributed is required for this approximation. However, it is worth noting that in practice, the structural shocks are unlikely to be normally distributed, nor do they necessarily have zero mean. Moreover, it is likely that $corr(u_t, u_{t'}) \neq 0$ for $t \neq t'$, so that it can be argued that u_t should be modeled in a way reflecting this correlation over time, rather than assuming it away. Consequently, the findings generated by this approach should be interpreted with caution. Moreover, it should be noted that estimating the model from Equation (15) requires more advanced techniques than the models discussed earlier in this section. In particular, estimating and interpreting σ^2 and u_t is not straightforward. However, this could be outweighed by the substantial benefits in terms of flexibility.

3.2.4 PD Model with the State-of-the-Economy as a GAS Recursion

Although the MEL model does allow for some flexibility in modeling u_t, due to the normality of u_t the MEL model is still somewhat restrictive. In order to model u_t in an alternative fashion, one might apply the Generalized Autoregressive Scoring (GAS) model defined by Creal et al. (2013). GAS models are a specific type of recursion to estimate models with time-varying parameters. In the application in this thesis, there is only one such time-varying parameter, being u_t. The main advantage of this model lies in the fact that it can model u_t and adjust the model fit based on observed quantities. Moreover, the requirement of collecting and selecting macroeconomic data is eliminated, which is desirable for the same reasons as outlined in the MEL model discussion. Similar to before, the starting point is the logistic model given by

$$logit(p^j_{t,t+12}) = Log\left(\frac{p^j_{t,t+12}}{1-p^j_{t,t+12}}\right) = \beta_0 + \sum_{i=1}^{k}\beta_i X^j_{i,t} + u_t \; with \; u_t \sim GAS(p,q), \tag{16}$$

where $GAS(p,q)$ denotes the GAS equation which will be introduced later in this subsection. In Blasques et al. (2014) a GAS application to credit risk data is demonstrated. This thesis shall aim to provide a modification of the method proposed there. The GAS model will be used to model u_t, so in order to model the state-of-the-economy proxy affecting the default probabilities as a time series.

The theoretical framework required to define the GAS model will be provided first, followed by its application in the PD modeling context. To start, the vector of time-varying parameters u_t is not observed. Next, it is assumed that all D^j_{t+12} are generated from an underlying logistic distribution function as in

$$D^j_{t+12} \sim p\left(D^j_{t+12} \mid \mathcal{F}_t;\theta\right) = \begin{cases} 1-p^j_{t,t+12} = \frac{1}{1+exp(\beta_0+\sum_{i=1}^k \beta_i X^j_{i,t}+u_t)}, & \text{if } D^j_{t+12}=0 \\ p^j_{t,t+12} = \frac{exp(\beta_0+\sum_{i=1}^k \beta_i X^j_{i,t}+u_t)}{1+exp(\beta_0+\sum_{i=1}^k \beta_i X^j_{i,t}+u_t)}, & \text{if } D^j_{t+12}=1. \end{cases} \tag{17}$$

Here, θ is the set of the model parameters to be estimated. This set includes β as well as the GAS parameters which will be introduced later in this subsection. Information matrix $\mathcal{F}_t$ contains information about all past default statuses, loan characteristics and states-of-the-economy. Therefore, define $\mathcal{F}_t = \left\{\mathbf{D}^j_t, \mathbf{X}^j_t, \mathbf{U}^t\right\}$ where the individual components are given by the matrices and vectors

$$\mathbf{D}^j_t = \begin{bmatrix} D^j_1 \\ D^j_2 \\ \vdots \\ D^j_t \end{bmatrix}, \mathbf{X}^j_t = \begin{bmatrix} X^j_{1,1} & X^j_{2,1} & \dots & X^j_{k,1} \\ X^j_{1,2} & X^j_{2,2} & \dots & X^j_{k,2} \\ \vdots & \vdots & \ddots & \vdots \\ X^j_{1,t} & X^j_{2,t} & \dots & X^j_{k,t} \end{bmatrix} \text{ and } \mathbf{U}^t = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_t \end{bmatrix}. \tag{18}$$

The idea is now that u_t follows a recursive relation based on the score of the observations and past realizations of u_t. The score is the derivative of the likelihood of the current observations with respect to a given set of parameters. Accordingly, the score with respect to u_t is defined by

$$\nabla_{j,t} = \frac{\partial \ln p\left(D^j_{t+12} \mid \mathcal{F}_t;\theta\right)}{\partial u_t}. \tag{19}$$

In Blasques et al. (2014), it is shown that in the panel setting the gradient ∇_t can be written as a sum of the scores over all mortgages in L_t, which is the set of mortgages active at time t. Write

$$\nabla_t = \sum_{j\in L_t} \nabla_{j,t} = \sum_{j\in L_t} \frac{\partial \log p\left(D^j_{t+12} \mid F_t;\theta\right)}{\partial u_t}. \tag{20}$$

This score is then incorporated in the GAS equation for u_t. The GAS(p,q) equation is then given by

$$u_{t+1} = \omega + \sum_{i=1}^{p} A_i s_{t-i+1} + \sum_{j=1}^{q} B_j u_{t-j+1}, \text{ where } s_t = S_t \cdot \nabla_t,, \quad S_t = S\left(\mathcal{F}_t;\theta\right), \tag{21}$$

where ω is a vector of constants, coefficient matrices A_i and B_j have appropriate dimensions for $i = 1,\dots,p$ and $j = 1,\dots,q$ and S_t is a scaling function of the score. All parameters $\{u_0,\ \omega,\ A_i,\ B_j\} \in \theta$ for $i = 1,...,p$ and $j = 1,...,q$ are to be estimated.

Assuming that the functional form of $p\left(D^j_{t+12} \mid \mathcal{F}_t;\theta\right)$ is given by Equation (17), the expression for ∇_t can easily be derived. The full derivation of ∇_t can be found in Appendix B.5.2. After some rewriting, it becomes clear that for the distribution function assumed in this thesis, ∇_t can be written as

$$\nabla_t = \sum_{j\in L_t} D^j_{t+12} - \frac{e^{\beta_0+\sum_{i=1}^k \beta_i X^j_{i,t}+u_t}}{e^{\beta_0+\sum_{i=1}^k \beta_i X^j_{i,t}+u_t}+1}. \tag{22}$$

Due to the complex nature of this approach, it might be good to take a step back and consider how to interpret this relation. From Equation (21), it becomes clear that ∇_t can be regarded as a sort of gradient in which direction u_t should move to reflect the observations in the data.

What is left to be defined is the scaling function $S_t\left(\mathcal{F}_t;\theta\right)$. Typically, the Fisher information matrix given by $\mathcal{I}_{t|t-1}^{-d}$ with $d\in\{0,\frac{1}{2},1\}$ is used as the scaling function $S(\cdot)$. The Fisher information matrix is defined as

$$S_t=\mathcal{I}_{t-1}^{-d}=-E_{t-1}\left[\frac{\partial^2\ln p\left(D_{t+12}^j\mid\mathcal{F}_t;\theta\right)}{\partial^2 u_t}\right]^{-d}. \tag{23}$$

In Appendix B.5.3 it is shown that for the application in this thesis, the following relation holds

$$S_t=I_{t|t-1}^{-d}=E_{t-1}\left[\frac{e^{\beta_0+\sum_{j=1}^{k}\beta_i X_{i,t}^j+u_t}}{\left(e^{\beta_0+\sum_{j=1}^{k}\beta_i X_{i,t}^j+u_t}+1\right)^2}\right]^{-d}. \tag{24}$$

Throughout this thesis p and q are assumed to be equal to one. Thus, a $GAS(1,1)$ setting is assumed. Moreover, staying consistent with the notation in the rest of this thesis, denote the next realization of u_t by u_{t+12} instead of u_{t+1}. Using this assumption and the definitions provided above, the updating equation becomes

$$u_{t+12}=\omega+A\cdot S_t\cdot\left[\sum_{j\in L_t}D_{t+12}^j-\frac{e^{\beta_0+\sum_{i=1}^{k}\beta_i X_{i,t}^j+u_t}}{e^{\beta_0+\sum_{i=1}^{k}\beta_i X_{i,t}^j+u_t}+1}\right]+Bu_t. \tag{25}$$

Throughout this thesis, it shall be assumed that $\mathcal{F}_t$ is simply $\mathcal{F}_t=\{D_t^j,X_{1,t}^j,X_{1,t}^j,\dots X_{k,t}^j,u_t\}$. That is, all past observations older than those in this information set do not affect the default-generating function from Equation (17). This is not required from the GAS methodology, but for this thesis it is sufficient and it makes notation slightly more intuitive. By the same argument, $\{X_{1,t}^j,X_{1,t}^j,\dots,X_{k,t}^j\}$ will be abbreviated to simply X_t^j.

If furthermore a value for d from Equation (24) is set, the maximum likelihood objective can be evaluated. The expressions required for this are provided in Section 4.3.

3.2.5 Summary of the PD Models and Assumptions

In Figure 3, an overview of the model assumptions for the PD models is provided.

PD Model Assumption Summary:

1. Observations: In each period t a set of mortgages L_t is observed containing a subset of all mortgages $j = 1, ..., J$ in the sample. The set mortgages from past periods are known and this set is denoted $\mathcal{L}_t = \{L_1, \dots, L_t\}$ for periods $t = 1, \dots, T$.
2. Panel Structure: The data form an unbalanced panel where a mortgage j observed at time t might or might not be observed at $t' \neq t$.
3. Observability: For each mortgage j a default indicator D_t^j and loan characteristics X_t^j are observed.
4. Default Condition: The realization of D_t^j is modelled by a latent default index D_{t+12}^{j*} as follows $D_{t+12}^j = I(D_{t+12}^{j*} > 0) = I\left(\beta_0 + \sum_{i=1}^k \beta_i X_{i,t}^j + u_t + \varepsilon_{jt} > 0\right)$ for all $j = 1, ..., J$ and $t = 1, ..., T$.
5. Distribution of Error: Idiosyncratic error ε_{jt} follows an i.i.d. logistic distribution $\varepsilon_{jt} \sim \text{Logistic}(0,1)$ and is independent of X_{it}^j for all $i = 1, ...k$, $j = 1, ..., J$, $t = 1, ..., T$ and is also independent of u_t for all $t = 1, ..., T$
6. The h-month probability of default is defined as $p_{t,t+h}^j = \mathbb{P}(D_{t+h}^j = 1 \mid X_t^j, u_t, D_t^j = 0) = E(D_{j,t+h} \mid X_t^j, u_t, D_t^j = 0)$ with horizon $h = 12$.
7. Absorbing Default State: $\mathbb{P}(D_{t+h}^j = 1 \mid D_t^j = 1) = 1$ for all $j = 1, ..., J$ and all $h \geq 0$.

---

8. Systematic Error: The error term u_t is modeled independently for each $t = 1, \dots, T$ and is independent of X_t^j and ε_{jt}.
   8.1. Benchmark Model: $u_t = 0$ for all $t = 1, ..., T$.
   8.2. Macroeconomic Model: $u_t = \sum_{i=1}^{k_M} \gamma_i M_{it}$ for all $t = 1, ..., T$ where M_{it} is an observed macroeconomic quantity.
   8.3. MEL Model: $u_t \sim N(0, \sigma^2)$ i.i.d. for all $t = 1, ..., T$.
   8.4. GAS: $u_t \sim GAS(1,1)$ for all $t = 1, ..., T$.

Figure 3: *Listing of the Model Assumptions for the PD Estimates and the Various Approaches to Modeling the State-of-the-Economy*

3.3 Alternative PD Models

One important point to note concerns the form of $p_{t,t+12}^j$. Throughout this thesis, it is assumed that the probability of default follows a logistic distribution. This choice is common practice in the credit risk management industry, as also mentioned in Duffie & Singleton (2012). However, there are alternative methods for estimating the probability of default, such as Survival Analysis (Cao et al., 2009), Random Forest (Mageto et al., 2015), Gradient Boosting Algorithms (Xia et al., 2021), and Markov chain Modeling (Kiefer & Larson, 2014). Although these alternatives are not discussed in further detail here to maintain consistency, it is important to be aware of their existence.

4 Estimation

In this section, it will be discussed how to estimate β, u_t and the default probabilities for the models that were described in the previous section. To do so, the maximum likelihood objectives for the models are provided. The estimation process will be similar across all models, although the estimation of u_t requires some additional steps in the MEL and GAS models.

The estimation procedures are all defined and carried out using *PYTHON* using a combination of user-defined functions and various libraries for the more standard computations. These include the Pandas (McKinney, 2023) library for data manipulation and analysis, the Numpy (Oliphant, 2023) library for numerical operations, and the sklearn (Cournapeau et al., 2023) library for machine learning and data processing.

4.1 Maximum Likelihood Estimation of the Benchmark PD Model

The starting point to derive the likelihood expressions is that outcome variable D^j_{t+12} is binary. The estimation target is the probability that this indicator equals 1 conditional on loan-specific characteristics X^j_t and state-of-the-economy u_t. The target quantity to be estimated throughout this thesis is

$$p^j_{t,t+12} = \mathbb{P}(D^j_{t+12} = 1 \mid u_t, X^j_t, D^j_t = 0). \tag{26}$$

For simplicity, assume that u_t is zero for now. To find the maximum likelihood expression for the equation above start from the the expression of the likelihood of binary outcome models presented in Cameron & Trivedi (2005). This can be altered slightly to resemble the notation used in this thesis to obtain

$$L(\beta \mid X_{\mathcal{T}}, D_{\mathcal{T}}) = \prod_{t\in\mathcal{T}} \prod_{j\in L_t} \mathbb{P}\left(D^j_{t+12} = 1 \mid X^j_t\right)^{D^j_{t+12}} \cdot \left(1 - \mathbb{P}\left(D^j_{t+12} = 1 \mid X^j_t\right)\right)^{1-D^j_{t+12}}. \tag{27}$$

Where the set $\mathcal{T}$ denotes the set of periods of interest and L_t the set of loans active in period t. The set $D_{\mathcal{T}} = \{D^j_{t+12} \mid j \in L_t, t \in \mathcal{T}\}$ denotes the set of the observed default indicators of all loans in period t and similarly $X_{\mathcal{T}} = \{X^j_t \mid j \in L_t, t \in \mathcal{T}\}$ denotes the set of observed loan specific characteristics of all loans observed in period t. Since it is assumed that

$$\mathbb{P}\left(D^j_{t+12} = 1 \mid u_t, X^j_t, D^j_t = 0\right) = \log\left(\frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^k \beta_i X^j_{i,t})}}\right), \tag{28}$$

the likelihood objective can be readily obtained. After calculating the logarithms of Equation (28) and substituting these in Equation (27), the log-likelihood function is obtained

$$\log L(\beta \mid X_{\mathcal{T}}, D_{\mathcal{T}}) = \sum_{t\in\mathcal{T}} \sum_{j\in L_t} \left[D^j_{t+12}(\beta_0 + \sum_{i=1}^k \beta_i X^j_{i,t}) - \log\left(1 + e^{(\beta_0 + \sum_{i=1}^k \beta_i X^j_{i,t})}\right)\right]. \tag{29}$$

Where some intermediate steps have been omitted for conciseness. For completeness' sake, these are provided in Appendix B.2. The optimization objective to obtain the estimates for β is then

$$\hat{\beta} = min_\beta - \sum_{t\in\mathcal{T}} \sum_{j\in L_t} \left[D^j_{t+12}(\beta_0 + \sum_{i=1}^k \beta_i X^j_{i,t}) - \log\left(1 + e^{(\beta_0 + \sum_{i=1}^k \beta_i X^j_{i,t})}\right)\right]. \tag{30}$$

This can then be solved through an optimization algorithm. Throughout this thesis *L-BFGS-B* (Zhu et al., 1997) shall be used to this extent.[4] Using $\hat{\beta}$, the 12-month PD can then be estimated via

$$\hat{p}^j_{t,t+12} = \frac{exp(\hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i X^j_{i,t})}{1 + exp(\hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i X^j_{i,t})} \tag{31}$$

4.1.1 Extension to PD Model with Explicit Macroeconomic Characteristics

For the model with macroeconomic characteristics, a nearly identical derivation can be performed by adding the extra $\sum_{i=1}^{k_M} \gamma_i M_{it}$ terms. This is omitted here, but can be found in Appendix B.3. The optimization objective is then

$$(\hat{\beta}, \hat{\gamma}) = min_{\beta,\gamma} - \sum_{t\in\mathcal{T}} \sum_{j\in L_t} \left[D^j_{t+12}(\beta_0 + \sum_{i=1}^k \beta_i X^j_{i,t} + \sum_{i=1}^{k_M} \gamma_i M_{it}) - \log\left(1 + e^{(\beta_0 + \sum_{i=1}^k \beta_i X^j_{i,t}) + \sum_{i=1}^{k_M} \gamma_i M_{it}}\right)\right]. \tag{32}$$

And similarly, the estimated default probabilities become

$$\hat{p}^j_{t,t+12} = \frac{exp(\hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i X^j_{i,t} + \sum_{i=1}^{k_M} \hat{\gamma}_i M_{it})}{1 + exp(\hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i X^j_{i,t} + \sum_{i=1}^{k_M} \hat{\gamma}_i M_{it})}. \tag{33}$$

[4] Also the BFGS and Nelder-Mead optimization algorithms were considered during the research. However, due to issues with computation time with the former and convergence with the latter, it was decided to perform all likelihood optimizations with L-BFGS-B.

4.2 Maximum Likelihood Estimation of the Mixed Effects Logistic PD Model

To estimate the MEL PD model defined by

$$logit(p^j_{t,t+12}) = Log\left(\frac{p^j_{t,t+12}}{1-p^j_{t,t+12}}\right) = \beta_0 + \sum_{i=1}^{k}\beta_i X^j_{i,t} + u_t \text{ with } u_t \sim N(0,\sigma^2) \text{ i.i.d. and unobserved,} \quad (34)$$

Demidenko (2013) and Mcculloch & Neuhaus (2013) provide detailed explanations. These will be used to estimate Equation (34). A summary of this estimation procedure based on these texts is provided in this section.

To find the parameter estimates for Equation (34), Maximum Likelihood Estimation is performed. However, due to the randomness of u_t, some additional steps are required. In particular, one can proceed by performing a two-step maximum likelihood estimation where in the first step u_t is integrated out. For the model provided in Equation (34) the likelihood conditional on the realizations of u_t is

$$\log L(\beta,\sigma \mid X_{\mathcal{T}}, D_{\mathcal{T}}, u_t) = \sum_{t\in\mathcal{T}}\sum_{j\in L_t}\left[D^j_{t+12}(\beta_0 + \sum_{i=1}^{k}\beta_i X^j_{i,t} + u_t) - \log\left(1+e^{\beta_0+\sum_{i=1}^{k}\beta_i X^j_{i,t}+u_t}\right)\right]. \quad (35)$$

Because the realization of u_t is unknown, the unconditional likelihood should be considered in order to proceed. This can be done by integrating out the u_t from Equation (35). The estimate for β and σ can be retrieved by minimizing the equation below

$$\log L(\beta,\sigma \mid X_{\mathcal{T}}, D_{\mathcal{T}}) = \sum_{t\in\mathcal{T}}\sum_{j\in L_t}\left[\int_{-\infty}^{\infty} D^j_{t+12}(\beta_0 + \sum_{i=1}^{k}\beta_i X^j_{i,t} + u_t) - \log\left(1+e^{\beta_0+\sum_{i=1}^{k}\beta_i X^j_{i,t}+u_t}\right)\ f_u(u)du\right]. \quad (36)$$

In order to estimate the integral in Equation (36), Gauss-Hermite quadrature will be applied. This is a technique to numerically approximate integrals by evaluating that integral at a moderate amount of nodes and calculating a weighted sum at the various nodes. Throughout this thesis, 25 nodes will be used for this approximation. A brief discussion of Gauss-Hermite quadrature is provided in Appendix A. The rewriting of Equation (36) to fit the form required for Gauss-Hermite quadrature is provided in Appendix B.4. After the rewriting, Equation (36) is approximated by

$$\log L(\beta,\sigma \mid X_{\mathcal{T}}, D_{\mathcal{T}}) \approx \sum_{t\in\mathcal{T}}\sum_{j\in L_t}\left[\sum_{h=1}^{H}\frac{w_h}{\sqrt{\pi}}(D^j_{t+12}(\beta_0 + \sum_{i=1}^{k}\beta_i X^j_{i,t} + \sqrt{2\sigma^2}z_h) - \log\left(1+e^{\beta_0+\sum_{i=1}^{k}\beta_i X^j_{i,t}+\sqrt{2\sigma^2}z_h}\right)\right]. \quad (37)$$

Where the first H nodes from the Hermite Polynomial are used and w_{h_n} are the corresponding weights, (see also Appendix A). To proceed, minimize this objective with respect to σ and β to obtain the estimates for these parameters. Write

$$(\hat\beta,\hat\sigma) = \min_{\beta,\sigma} - \sum_{t\in\mathcal{T}}\sum_{j\in L_t}\left[\sum_{h_n=1}^{H}\frac{w_{h_n}}{\sqrt{\pi}}D^j_{t+12}(\beta_0 + \sum_{i=1}^{k}\beta_i X^j_{i,t} + \sqrt{2\sigma^2}z_{h_n}) - \log\left(1+e^{\beta_0+\sum_{i=1}^{k}\beta_i X^j_{i,t}+\sqrt{2\sigma^2}z_{h_n}}\right)\right]. \quad (38)$$

If one is not interested in the realizations of u_t, but only in the values of $\hat\beta$ and $\hat\sigma$, one could in principle stop here. In the results section, the value of the likelihood function in Equation (38) will serve as input for the test statistic in determining whether σ is significantly different from zero.

In this thesis, the primary goal is to estimate the values of u_t, rather than the value of its standard error σ. To retrieve the values of u_t, one can maximize likelihood function $\log L(u_t \mid \hat\beta,\hat\sigma,X_{\mathcal{T}},D_{\mathcal{T}})$ as below

$$\begin{aligned}\hat u^*_t &= min_{u_t} - \log L(u_t \mid \hat\beta,\hat\sigma,X_{\mathcal{T}},D_{\mathcal{T}}) \\ &= min_{u_t} - \sum_{t\in\mathcal{T}}\sum_{j\in L_t}\left[D^j_{t+12}(\hat\beta_0 + \sum_{i=1}^{k}\hat\beta_i X^j_{i,t} + u_t) - \log\left(1+e^{\hat\beta_0+\sum_{i=1}^{k}\hat\beta_i X^j_{i,t}+u_t}\right)\right].\end{aligned} \quad (39)$$

However, by defining $\hat u^*_t$ as in Equation (39), near equality between the average of estimated PDs within a given period and Observed Default Rates within that same period arises[5]. To address this issue, the estimate $\hat u^*_t$ is replaced with $\hat u^*_{t-12}$. This modification diminishes reliance on forward-looking data.

[5]This issue is not due to collinearity; eliminating the intercept or certain year dummies does not resolve it.

Additionally, as will be discussed in Section 6, this issue is exclusively relevant when PDs are estimated based on other loans within the same period. When PD estimates are constructed based on loans observed in the past, the estimate in Equation (39) is not possible to construct in the first place. Both methods of constructing the PDs will be assessed later. Thus, in addition to resolving the reliance on forward-looking data, shifting the estimate $\hat{u}_t^*$ also facilitates the comparison across different applications of the model. To summarize, define the estimate of $\hat{u}_t^*$ by

$$\hat{u}_t = \hat{u}_{t-12}^*. \tag{40}$$

To then construct estimates for the default probability, again calculate

$$\hat{p}_{t,t+12}^j = \frac{exp(\hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i X_{i,t}^j + \hat{u}_t)}{1 + exp(\hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i X_{i,t}^j + \hat{u}_t)}. \tag{41}$$

4.3 Maximum Likelihood Estimation of the PD Model with the State-of-the-Economy as a GAS Recursion

In order to derive the likelihood expression for the PD model where u_t follows a GAS recursion, recall that

$$logit(p_{t,t+12}^j) = Log\left(\frac{p_{t,t+12}^j}{1 - p_{t,t+12}^j}\right) = \beta_0 + \sum_{i=1}^k \beta_i X_{i,t}^j + u_t \text{ with } u_t \sim GAS(1,1). \tag{42}$$

The estimation objective is in this case

$$\hat{\theta} = arg\ max_\theta \sum_{t\in\mathcal{T}} \sum_{t=1}^n ln\ p\left(D_{t+12}^j \mid \mathcal{F}_t; \theta\right). \tag{43}$$

Where $\theta = \{\beta, u_0, \omega, A, B\}$ and $\mathcal{F}_t = \{D_t^j, X_t^j, u_t\}$. The GAS parameters $\{u_0, \omega, A, B\}$ will be introduced later in this section. One can estimate Equation (43) to obtain estimates for the coefficients of the risk drivers. Since D_{t+12}^j is assumed to be generated by probability density function $p\left(D_{t+12}^j \mid \mathcal{F}_t; \theta\right)$ this can be written as

$$D_{t+12}^j \sim p\left(D_{t+12}^j \mid \mathcal{F}_t; \theta\right) = \begin{cases} 1 - p_{t,t+12}^j = \frac{1}{1+exp(\beta_0 + \sum_{i=1}^k \beta_i X_{i,t}^j + u_t)}, & \text{if } D_{t+12}^j = 0 \\ p_{t,t+12}^j = \frac{exp(\beta_0 + \sum_{i=1}^k \beta_i X_{i,t}^j + u_t)}{1+exp(\beta_0 + \sum_{i=1}^k \beta_i X_{i,t}^j + u_t)}, & \text{if } D_{t+12}^j = 1. \end{cases} \tag{44}$$

Equivalently to before, in Appendix B.5.1 it is shown that this expression can be rewritten into

$$\begin{aligned} \hat{\theta} &= min_\theta - \sum_{t\in\mathcal{T}} \sum_{j\in L_t} \left[D_{t+12}^j (\beta_0 + \sum_{i=1}^k \beta_i X_{i,t}^j + u_t) - \log\left(1 + e^{\beta_0 + \sum_{i=1}^k \beta_i X_{i,t}^j + u_t}\right) \right], \text{ where} \\ u_{t+12} &= \omega + A \cdot S_t \cdot \left[\sum_{j\in L_t} D_{t+12}^j - \frac{e^{\beta_0 + \sum_{i=1}^k \beta_i X_{i,t}^j + u_t}}{e^{\beta_0 + \sum_{i=1}^k \beta_i X_{i,t}^j + u_t} + 1} \right] + B u_t. \end{aligned} \tag{45}$$

If then an estimate for the scaling function $\hat{S}_t$ is provided, the estimates $\hat{\theta}$ can be constructed. Recall from Equation (24) that scaling function S_t is defined by

$$S_t = I_{t|t-1}^{-d} = -E_{t-1}\left[- \frac{e^{\beta_0 + \sum_{i=1}^k \beta_i X_{i,t}^j + u_t}}{\left(e^{\beta_0 + \sum_{i=1}^k \beta_i X_{i,t}^j + u_t} + 1\right)^2} \right]^{-d}. \tag{46}$$

However, since estimating the expectation of this expression given the information at time $t-1$ is challenging, and the main purpose of this quantity is to ensure computational stability, the subscript is omitted. Moreover, d is set $d = 1$. The final functional form of S_t used in the estimation algorithm then becomes

$$S_t = -E\left[- \frac{e^{\beta_0 + \sum_{i=1}^k \beta_i X_{i,t}^j + u_t}}{\left(e^{\beta_0 + \sum_{i=1}^k \beta_i X_{i,t}^j + u_t} + 1\right)^2} \right]^{-1} = E\left[\frac{e^{\beta_0 + \sum_{i=1}^k \beta_i X_{i,t}^j + u_t}}{\left(e^{\beta_0 + \sum_{i=1}^k \beta_i X_{i,t}^j + u_t} + 1\right)^2} \right]^{-1}. \tag{47}$$

Which is estimated by

$$\hat{S}_t = \frac{1}{|L_t|} \sum_{j \in L_t} \left[\frac{e^{\hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i X_{i,t}^j + \hat{u}_t}}{\left(e^{\hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i X_{i,t}^j + \hat{u}_t} + 1\right)^2} \right]^{-1}. \tag{48}$$

If Equation (45) is then solved for θ, the 12-month Point-in-Time PD can be obtained by

$$\hat{p}_{t,t+12}^j = \frac{exp(\hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i X_{i,t}^j + \hat{u}_t)}{1 + exp(\hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i X_{i,t}^j + \hat{u}_t)}. \tag{49}$$

4.4 Parameter Selection Procedure

In order to prevent overfitting, it is important to carefully select the parameters used in the analysis. This is particularly important because there is a substantial number of parameters. In order to achieve this, two variable selection methods will be used. These will be stepwise regression by means of forward selection based on Akaike's Information Criterion (AIC) (Akaike, 1973) and stepwise regression by means of backward selection based on the AIC. These procedures may not necessarily give the same resulting set of risk drivers. So if a risk driver is not present in either of the two sets or only present in one of the two sets, it will be decided to not use the risk driver in further analysis.

Akaike's Information Criterion (AIC) is a measure for comparing the likelihood of a model against its number of parameters, offering a way to compare different models to find a balance between model fit and model complexity. The AIC value is calculated as

$$AIC = -2 \cdot \log(L) + 2 \cdot k, \tag{50}$$

where L is the maximized likelihood of the model and k is the number of parameters. The model with the lowest AIC value is considered the best balanced.

For forward selection based on the AIC, the procedure to select the parameters to include in the estimation of the PD will be to use step-wise regression starting at $k = 1$ which is the intercept-only model and then iteratively include an additional parameter such that the increase of the AIC is maximal. That is, the least negative. This is repeated until the AIC no longer decreases. The stepwise regression by means of forward selection is outlined in Appendix D.1.

Backward stepwise regression will start by selecting all risk drivers including the intercept. In each iteration, the risk driver without which the model likelihood is smallest is the excluded parameter. This is repeated until the AIC no longer decreases. Stepwise regression by means of backward selection is outlined in Appendix D.2.

4.5 Performance Measurement

In credit risk management, PD model performance evaluation is an essential process that generally uses two metrics. These are calibration accuracy and discriminatory power. Both are described in this subsection. The Mean Squared Error (MSE), which measures the difference between estimated default probabilities and observed defaults, is used to assess the calibration accuracy. Conversely, the model's ability to discriminate between loans that will be in default and those that will not be is determined by the Area Under the ROC Curve, (AUC). These are both standard metrics often encountered in credit risk modeling, as noted by the BCBS (2005) and Medema et al. (2009). The subsections that follow shall be based on these sources.

4.5.1 Calibration Accuracy

Calibration accuracy evaluates the precision of the predicted outcome variables in comparison to the realized outcomes. In the case of this thesis these are $\hat{p}_{t,t+12}^j$ and D_{t+12}^j respectively. Essentially, the calibration accuracy of the PD models evaluates whether the predicted PDs accurately reflect the true PD. High calibration accuracy implies that the estimates are reliable.

The calibration accuracy of credit risk models can be assessed using various metrics. In this thesis, this shall be done using the MSE, which measures the accuracy of the model by quantifying the mean squared difference between predicted probabilities and actual outcomes. A lower MSE indicates a higher degree of calibration. In this thesis, two different MSE scores shall be computed for each model.

Firstly, overall performance will be assessed by comparing MSE between each predicted default probability of the individual mortgages and the actual outcomes. If the model is well-calibrated, mortgages with a higher estimated PD will default more often which will lead to a smaller estimation error. This "individual MSE" is defined as

$$MSE_{individual} = \frac{1}{|\mathcal{L}_{\mathcal{T}}|} \sum_{L_t \in \mathcal{L}_{\mathcal{T}}} \sum_{j \in L_t} (\hat{p}^j_{t,t+12} - D^j_{t+12})^2, \tag{51}$$

where $|\mathcal{L}_{\mathcal{T}}|$ represents the total number of observations of loans active in the set of periods of interest $\mathcal{T}$.

Secondly, the MSE of all observations in a given year shall be compared to the default rate in that year. The purpose of this is that it allows for the measurement of the accuracy of the aggregate default rate in a given year, which assesses to what extent a model captures the fluctuations in the overall default rate more clearly than the individual MSE. This MSE comparing the estimates and realizations within a given year is defined as

$$MSE_{Annual} = \frac{1}{|\mathcal{L}_{\mathcal{T}}|} \sum_{t \in \mathcal{T}} |L_t| \cdot (\bar{\hat{p}}_{t,t+12} - \bar{D}_{t+12})^2, \tag{52}$$

where $\bar{D}_{t+12} = \frac{1}{|\mathcal{L}_t|} \sum_{j \in L_t} D^j_{t+12}$ and $\bar{\hat{p}}_{t,t+12} = \frac{1}{|\mathcal{L}_t|} \sum_{j \in L_t} \hat{p}^j_{t,t+12}$.

4.5.2 Discriminatory Power

Discriminatory power is the ability of a model to discriminate between positive and negative outcomes. In the context of PD estimation, this amounts to discriminating between defaults and non-defaults. This classification can be done by introducing a threshold p^*. Using this threshold, an estimate for default status $\hat{D}^j_{t+12}$ can be constructed such that $\hat{D}^j_{t+12} = \mathbb{1} * (\hat{p}^j_{t,t+12} \geq p^*)$. Here, $\mathbb{1}$ is an indicator function that equals 1 when the condition $\hat{p}^j_{t,t+12} \geq p^*$ is met (indicating a predicted default) and 0 otherwise. Using this threshold, one can construct an estimate for default status $\hat{D}^j_{t+12}$ such that $\hat{D}^j_{t+12} = \mathbb{1} * (\hat{p}^j_{t,t+12} \geq p^*)$. That is, if the predicted default probability $\hat{p}^j_{t,t+12}$ is larger than or equal to p^*, one would classify this mortgage as defaulting.

Although in this thesis these notions shall not directly be applied nor will the determination of individual defaults via a classification threshold be pursued, the principle of classification allows for the construction of the ROC curve. The ROC curve can in turn be used to calculate the AUC, the measure of discriminatory power that will be used in this thesis.

The ROC curve plots the true positive rate against the false positive rate for different values of p^*. To construct the ROC curve, one can compute both the true positive rate and the false positive rate for every threshold p^* in the range $0 \leq p^* \leq 1$. Define the relative frequencies of the True Positives (TP) and False Positives (FP) as

$$TP(p^*) = \frac{1}{|L_{\mathcal{T}}|} \sum_{j \in L_{\mathcal{T}}} \mathbb{1} * (\hat{D}^j_{t+12} = D^j_{t+12} = 1) \text{ and } FP(p^*) = \frac{1}{|L_{\mathcal{T}}|} \sum_{j \in L_t} \mathbb{1} * (\hat{D}^j_{t+12} = 1 \neq D^j_{t+12}), \tag{53}$$

where $L_{\mathcal{T}}$ is the set of mortgages in the set of periods $\mathcal{T}$ and $|L_{\mathcal{T}}|$ is the total number of observations in this set. An example of this ROC curve is shown in Figure 4.

The AUC is then the area under the ROC curve. A larger AUC value signifies a higher degree of discriminatory power. To calculate the AUC from the ROC, the area under the ROC curve on the interval $0 \leq p^* \leq 1$ is calculated[6].

[6] In some applications, $\frac{1}{2}$ is subtracted from this area. However, in this thesis, the AUC is only used to "rank" the models and if all models are shifted by the same amount, the "ranking" from largest to smallest AUC remains the same. For simplification, this subtraction is therefore omitted.

To numerically approximate the area under the ROC curve, one can use a Riemann sum and determine the true positive rate for k different values of the cut-off threshold p^* and subtract the area from the random model. Then, set the bandwidth of the Riemann sum to $h = \frac{1}{k}$. Write

$$AUC = \sum_{i=0}^{i=k-1} TP(i * h) * h. \tag{54}$$

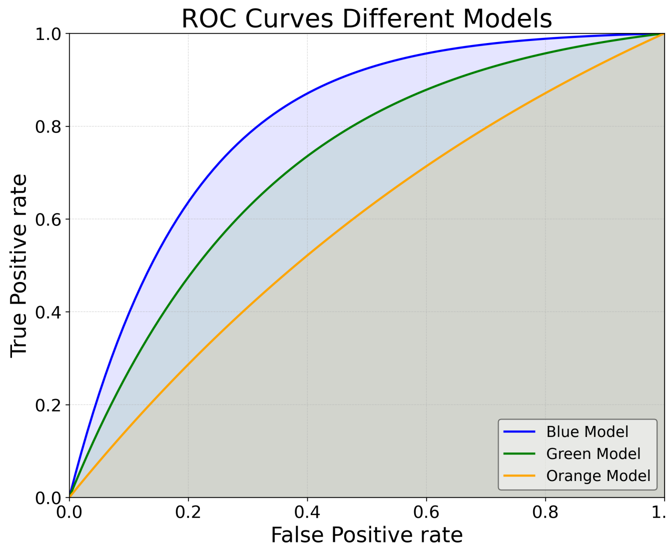


Figure 4: *ROC curve comparison of three models. The shaded area is the AUC for each of the models. The larger the area of the model, the larger the degree of discriminatory power. In this figure, the "blue model" thus has the largest degree of discriminatory power.*

4.5.3 Likelihood Ratio Test

Additionally, likelihood ratio tests will be conducted. The discussion in this subsection will be based on Chapter 12.8 of Bain & Engelhardt (1992). For this likelihood ratio test, it is required that one of the models is nested in the other. A nested model is a model which is the simpler version of a more complex model, usually with fewer parameters or restrictions. Typically, this involves the simpler model having certain parameters set to zero, which may be non-zero in the more complex model. In the application in this thesis, the benchmark model from Section 3.2.1 is a nested model for the models that do include a systematic error term. Specifically, in the benchmark model, u_t is assumed to be equal to zero, in contrast to the other models. The test statistic is of the likelihood ratio test then involves assessing the difference in likelihoods between the nested and the more complex and is given by

$$\lambda_{\text{LR}} = -2\left[\ell\left(\theta_0\right) - \ell(\hat{\theta})\right] \sim \chi^2(k - k_0). \tag{55}$$

With k_0 the number of parameters of the nested model and $\ell\left(\theta_0\right)$ the log-likelihood of the nested model evaluated at θ_0. Similarly, k is the number of parameters of the more complex model and $\ell\left(\theta\right)$ the log-likelihood of that model evaluated at $\hat{\theta}$. The test can then be rejected or failed to be rejected depending on whether λ_{LR} exceeds the relevant confidence threshold of the χ^2 with $k - k_0$ degrees of freedom.

5 Data Description

5.1 Mortgage Data

The consumer mortgage data is retrieved from the Federal Home Loan Mortgage Corporation, commonly known as Freddie Mac. Specifically, the single-family loan-level data set is used. Freddie Mac is a US government-sponsored enterprise that buys mortgages, pools them, and then sells these pooled mortgages as mortgage-backed securities to investors. Moreover, as mentioned on their own website[7], Freddie Mac

[7]https://www.freddiemac.com/research/datasets

makes their loan-level credit performance data available on a portion of mortgages purchased by the company. This is done with the aim of helping investors build their portfolios. Of course, this also opens the door for academic studies on the same data set, as will be done in this thesis.

The data that will be used spans from the year 2000 up to and including 2022. These data contain information about the default status of specific mortgages and additional information such as the loan to value, original interest rate and loan age. The data set is split into two smaller data sets. The first of these is an origination file, which contains information about the mortgage and the issuer at the time of issuing. The second is a performance file, which contains monthly updated information about the current status of the loan and the borrower. In both data sets, each observation also contains a "loan sequence number" which is a string uniquely assigned to each mortgage such that it is possible to track the entire history of a given loan in the data set.

Furthermore, a comment should be made regarding the size of the complete single-family data set. As of December 31st, 2022, the performance file comprises 2.4 billion observations. To be able to process the data and estimate the models in a reasonable amount of time, a sample is used instead. Freddie Mac offers such a sampled data set on its website. In that data set, a sample of mortgages was followed from 2000 through 2022 where 50,000 loans are randomly chosen and included in the data set each year. This is the data set that will be examined in the analysis. The assumption made in this thesis is that this sampled data set has the same general characteristics as the complete data set. This assumption is also supported by Freddie Mac. In their user guide, it is explained that their sampling method aims to ensure that each member of the subset has an equal chance of being selected from the larger population. This approach, known as a simple random sample, allows for an (asymptotically) unbiased representation of the broader population (Freddie Mac, 2023).

By default these mortgages are observed each month. To reduce memory requirements, only the observations occurring in the same month as the original payment are used, such that each mortgage is observed at most once each year. Moreover, since not all mortgages originate in the same month, the observations are spread out over the year by reducing the number of observations in this way.

In order to make a selection out of all risk drivers provided in the loan-level data set, first, all risk drivers where more than half of the observations are missing are removed from the data set. Thereafter, all categorical risk drivers with more than 10 categories were removed to reduce dimensionality. This comes down to removing data relating to the area where the property was located where removed to simplify estimation. For example, the "Postal Code" column consists of 892 unique values. A full description of all risk drivers in the original data set is not included in this thesis, but can be found in the single-family mortgage data set documentation by Freddie Mac (2023).

Thereafter, special attention is paid to the pairs: *Original UPB / Current UPB*, *Original Interest Rate/ Current Interest Rate*, *Original Loan To Value / Current Loan To Value* and *Original Debt To Income Ratio / Current Debt To Income Ratio*. These are all strongly correlated to each other. That is, all these pairs have an in-sample correlation of at least 0.8. To avoid any complications, it was decided to use the *Original* variants from each of these pairs and remove the *Current* variants in further analysis. Another pair from which such correlation might be expected is the pair *Original Loan To Value / Original Combined Loan To Value*, these have an in-sample correlation of 97%. Here, *Original Combined Loan to Value* is removed.

After removing these risk drivers, 15 risk drivers are remaining. All observations where any of the columns is missing shall be removed. In Table 1 an overview of these risk drivers is provided along with their description. In Appendix C.2, in Table 8 the summary statistics regarding the continuous risk drivers are reported and in Table 9 within the same appendix the relative frequencies observed within the categorical risk drivers are reported.

After creating dummy columns for the categorical variables, the risk drivers from Table 1 are used as candidate risk drivers in the variable selection procedure. The inclusion of each risk driver will be evaluated through both forward and backward selection. Additionally, an intercept will be considered as part of the set of risk drivers, as its inclusion could potentially enhance model performance. This procedure is only performed for the benchmark model. The variable selection for the model including macroeconomic risk drivers will start with the same set as the benchmark model. The macroeconomic risk drivers will then be used as new candidate risk drivers to potentially append this set with. For the MEL model and the GAS model, all parameters are appended to the set of variables used in the benchmark model.

Risk Driver	**Description**	**Data Type**
Credit Score	Number by third parties indicating borrower's creditworthiness	Numeric
Current Deferred UPB	Non-interest bearing UPB of the modified loan.	Numeric
Loan Age	Number of months since first mortgage payment	Numeric
Mortgage Insurance Percentage	The percentage of loss coverage on the loan, at the time of Freddie Mac's purchase of the mortgage loan that a mortgage insurer is providing to cover losses incurred as a result of a default on the loan	Numeric
Original Debt To Income Ratio	Monthly debt, including housing, divided by total monthly income at loan origination	Numeric
Original Interest Rate	Rate as shown on the mortgage note	Numeric
Original Loan To Value	Ratio of original mortgage to property's appraised value or purchase price	Numeric
Original UPB	Mortgage's unpaid balance on the note date	Numeric
Loan Term	Calculated number of scheduled monthly payments between the First Payment and Maturity Date	Numeric
First Time Homebuyer Flag	Indicates if a borrower is buying a primary home and had no property ownership in the past three years. Equals one if yes.	Categorical
Channel	Indicates origination method: Retail, Broker, Correspondent, or not specificity. Mortgages originating from a broker are the reference group.	Categorical
Loan Purpose	Indicates if the loan is for Cash-out Refinance, No Cash-out Refinance, or Purchase. Cash-out Refinanced mortgages are the reference group.	Categorical
Number Of Borrowers	Denotes if there is one borrower or more obligated to repay the mortgage note	Count
Number Of Units	Represents the number of units on the mortgaged property	Count

Table 1: *Mortgage-Specific Risk Driver Overview*

Risk Driver	Included / Excluded Forward Selection	Included / Excluded Backward Selection
Intercept	✓	✓
Credit Score	✓	✓
Current Deferred UPB	✓	✓
Loan Age	✓	✓
Mortgage Insurance Percentage	✓	✓
Original Debt To Income Ratio	✓	✓
Original Interest Rate	✓	✓
Original Loan To Value	✓	✓
Original UPB	✓	✓
Loan Term	✓	✓
Loan Purpose - No Cash-out Refinance	✓	✓
Loan Purpose - Purchase	✓	✓
First Time Homebuyer Flag - Yes	✓	✓
Channel - Correspondent	✓	✓
Channel - Retail	✓	✓
Number Of Borrowers	✓	✓
Number Of Units	×	×
Channel - Not specified	×	×

Table 2: *Inclusion status of risk drivers in the final data set.*

5.2 Definition of Default and Outcome Variable Construction

Multiple definitions can be used to determine if a mortgage is in default. In this subsection, the definition applied in this thesis will be described. The European Banking Authority (EBA), in Article 178 of the Capital Requirements Regulation (CRR) states that an obligor is in default when any material credit obligation is more than 90 days past due. Acknowledging the complexity of defining "material" in this context, this thesis will simply adopt the following definition of default

Definition 2) A mortgage is in default if a borrower is more than 90 days overdue on their payments.

In order to then construct outcome variable D^j_{t+12}, the *Loan Delinquency Status* column provided by Freddie Mac is used. This column provides a value of 0 if a loan payment is due for 0 to 30 days, a value of 1 for a payment due 30 to 60 days, a value of 2 for a payment due for 60 to 90 days and so forth. Therefore, if a loan 12 periods from now has a Loan Delinquency status of 3 or more, D^j_{t+12} will be set to 1 and 0 otherwise. As a consequence, for the last 12 months of the data, no estimate will be made for the probability of default. Therefore, the results will be reported until 2021 rather than until 2022.

In Appendix C.2.1 for each year, the total number of default D^j_{t+12} are shown in Figure 9 and Table 10.

5.3 Macroeconomic Data

5.3.1 Overview of Selected Macroeconomic Risk Drivers

For the macroeconomic data, five quantities on the US economy are considered and listed below, along with brief motivations for their inclusion. All of these quantities have been obtained from the website of the Federal Reserve Economic Data[8], commonly known as FRED. Each of these quantities is observed monthly, except for the Gross Domestic Product (GDP), which is observed quarterly.

Firstly, the US Unemployment rate is included. This rate indicates the percentage of the labor force that is without employment but is actively seeking work. This quantity is one of the most often used macroeconomic quantities for estimating mortgage PDs. Various studies, such as those by Foote et al. (2010), Foote et al. (2008), Pennington-Cross & Ho (2010), and Gerardi et al. (2007), have highlighted its correlation with mortgage defaults. The economic intuition for the relevance of employment status is straightforward: Employment status is closely related to an individual's income stability, which is a critical factor in their ability to fulfill financial commitments, such as mortgage repayments. Without

[8] https://fred.stlouisfed.org/

a reliable source of income, consistently making mortgage payments can become challenging. This difficulty is amplified if the individual lacks savings. In such circumstances meeting mortgage repayment obligations might become troublesome, thus increasing the risk of default.

Secondly, the GDP of the United States is included. The GDP is measured in billions of dollars and reported quarterly. Since this is the only quantity in this analysis measured quarterly, for all "missing" months, the most recently observed value is used to estimate the GDP. In Figlewski et al. (2006), Koopman & Lucas (2005b) and Simons & Rolwes (2009) the relation between mortgage defaults and the GDP is confirmed. The rationale behind using GDP as a predictor of mortgage defaults is that an increasing GDP often signals a growing economy. This economic growth directly influences the income levels and financial stability of homeowners. Higher income thereby reducing the likelihood of defaults. Of course, the effect of GDP on mortgage defaults is even broader and also some rationale from the opposite relation might apply. For instance, if housing prices increase in response to GDP growth, it might increase the size of mortgage repayments on average which in turn might increase mortgage default probability. That said, it is intuitive that there is at least some effect of GDP growth on mortgage default is clear and given the empirical evidence, this risk driver is included in the analysis.

Third, the US Consumer goods and services Price Index (CPI) will be included. It is often argued that substantial inflation rates adversely affect households' ability to meet their financial needs. This correlation has been substantiated in studies by for example Guo & Bruneau (2014) and Campbell & Cocco (2015). When inflation causes a general rise in prices at a rate that is faster than income growth, households might have fewer (real) financial resources, leaving less funds available for mortgage payments. In such scenarios, the strain on household finances can increase and thereby also the likelihood of mortgage defaults. Also, inflation is closely linked to monetary policy, notably in the form of interest rates which might be increased if inflation is high. Higher interest rates can lead to a rise in monthly mortgage payments. This increase in payment obligations places additional financial strain on homeowners. The effects of both the increased living expenses and the increased mortgage payments can reasonably be expected to affect the financial health of households.

Fourth, the CBOE® VIX, a volatility index measuring the volatility of the Standard & Poor's 500 market index, will be used. In Gavalas & Syriopoulos (2014) it is shown that the VIX affects the transition rate of mortgages from one category of credit ratings to another. It is reasonable to assume that the VIX thus also affects the mortgage default rate directly. Moreover, considering that the S&P 500 reflects at least part of the economic conditions of the US, it is reasonable to expect that increased volatility in the S&P 500 might stem from the same macroeconomic circumstances that affect the mortgage default rate. For instance, significant fluctuations in the S&P 500 can reflect investor uncertainties, shifts in market sentiment, or reactions to economic events, all of which are tied to the broader state-of-the-economy. These elements of economic uncertainty can impact the financial stability of households and thereby mortgage default probabilities.

The fifth macroeconomic risk driver that is considered is the Economic Policy Uncertainty Index (EPU) index. The EPU index is calculated using three components. The first component is the number of newspaper articles discussing economic policy uncertainty. The second component is the count of tax code provisions that are set to expire in the next 10 years. The third component measures the variation in economic forecasts, focusing on price indices and public expenditures. In this third component, greater variation indicates higher uncertainty. An extensive discussion on how this quantity is computed can be found in Baker et al. (2016). This index is perhaps somewhat obscure, but it does make sense to include it in this analysis since the target is to measure the effect of the state-of-the-economy on the probability of default. The EPU Index, by construction, captures the degree of uncertainty surrounding economic policies which can have implications on economic stability and predictability. Such uncertainties can influence consumer confidence, investment decisions and overall economic activity. These are all critical factors in the financial well-being of households and the general state-of-the-economy. Moreover, some empirical evidence of the relation between mortgage defaults and the EPU index is suggested by the findings in Chi & Li (2017).

5.3.2 Data Preprocessing

Each of the five considered quantities will be transformed to reflect its 12-month percentage change additional to the current absolute value. For each quantity, the correlation with the ODR is calculated, considering both its relative (percentage change) and absolute values, across various lags. For instance, in analyzing the unemployment rate, the current value's correlation and that of its lagged values (up to 24 months) with the ODR are calculated. Since the mortgage data are observed for the years 2000-2022, to generate the 24-month lagged 12-month relative change in unemployment, data from 1997-2022 is required. This is why all macroeconomic data are retrieved from that period. The intuition for including the lags is that a mortgage owner might not immediately fail to meet its repayment requirement the moment that, for example, the S&P 500 exhibits a large amount of volatility. Since it is unclear how long it might take and if this is correlated to the default rate at all, both the absolute and relative variants of the same quantity are considered up to a two-year lag[9].

After the evaluation of the variants of macroeconomic quantities, the variant exhibiting the largest correlation in absolute value with the default rate is selected for the variable selection procedure. Instead, they are used as candidate risk drivers to extend upon the existing benchmark model based on forward and backward selection.

In Appendix C.3.1 in Figure 10, the correlations between the macroeconomic quantities and loan default rates are reported. For the following configuration of the macroeconomic risk drivers, the correlations are maximized: a 12-month current percentage change for CPI, the current VIX in absolute terms, the EPU absolute lagged by 12 months, the absolute current unemployment rate, and the GDP percentage change lagged by 24 months. In Table 3, the inclusion status of each risk driver is reported.

To obtain some further idea of the values of the macroeconomic risk drivers over time, in Appendix C.3.2 in Figure 11, also the values of the macroeconomic quantities that are considered for the risk driver selection are plotted in the same figure as the Observed Default Rate.

Risk Driver	Included / Excluded Forward Selection	Included / Excluded Backward Selection
Consumer Price Index (Current Annual Change)	✓	✓
Volatility Index (Current Level)	✓	✓
Economic Policy Uncertainty Index (One-Year Lag)	✓	✓
Current Unemployment Rate	✓	✓
Gross Domestic Product (Two-Year Lag Annual Change)	✓	✓

Table 3: *Selection of Macroeconomic Variables After Forward and Backward Selection*

5.4 Estimation and Validation Process

This subsection will outline the methodologies employed to assess the performance of the predictive models developed in this thesis. The evaluation will be conducted through various train-test splits, where a subset of the data is utilized to train the models, and the remaining data is used to evaluate their performance. The results discussion in Section 6 will discuss three different splits: the Out-Of-Sample split, the one-year Out-Of-Time split and the three-year Out-Of-Time split. These will all be discussed below. Figure 5 summarizes the differences between the three splits visually. Based on the results within each of these splits, a comparison in model performance across the different settings will be made. Each of these individual results will be used to compile a single and nuanced answer to the research question in Section 7.

5.4.1 Train Test Splits: Out-Of-Sample Split

A common way to assess model performance is by performing a train test split. In such a split, part of a data set is sampled to provide the estimates for a model and the remaining sample is used to assess whether these estimates can accurately describe the outcome variable. In this thesis, this method shall be applied as well. To do so, the models will be fit on the training set which consists of 70% of the unique mortgages. The default probabilities obtained by the estimates of the testing set of the remaining 30%

[9]Also lags up to 60 months have been included in the analysis in earlier stages, but in none of the cases surveyed, the correlation was largest for lags larger than 24. Similarly, this procedure is also executed for 1-month relative and absolute changes as well as 12-month relative changes. However, the 12-month relative change or the current observed value consistently showed a larger correlation with the ODR, so only the cases with 12-month relative change and actual value are reported in this thesis.

are then compared to the realizations of the testing set. In the results section, this shall be referred to as the Out-Of-Sample split. The sampling is performed randomly where each loan has equal probabilities for being selected into either set. For each of the models, the training and the testing set will consist of the same observations to facilitate comparison.

5.4.2 Train Test Splits: Out-Of-Time Splits

Since the primary aim of this study is to develop a model for accurately predicting the 12-month PD at period t, also Out-Of-Time tests are conducted. In these tests, all available data up to a given point in time t is used to calibrate the models and then PDs in a later 12-month span are estimated. In this thesis, a one-year Out-Of-Time rolling forecast as well as three-year Out-Of-Time rolling forecast will be performed.

One-Year Out-Of-Time Rolling Forecast

To perform the one-year Out-Of-Time rolling forecast, models are fit using all historical data preceding the current period t. For example, to predict the PDs in 2011, models are trained using data from January 2000 through December 2010. For the subsequent year of which the PDs are estimated, 2012, the training data extends up to December 2011, thereby progressively incorporating additional historical information with each passing year. The objective here is to ensure the most accurate PD estimation possible by incorporating as much data as possible.

Three-Year Out-Of-Time Rolling Forecast

Additionally, a three-year Out-Of-Time forecast is introduced to reflect the recalibration requirements by the European Banking Authority (EBA), which do not require annual model updates. This longer forecast period aligns with the maximum recalibration horizon observed among financial institutions, as indicated in the EBA guidelines (EBA (2017), p. 147). Therefore, additionally, model performance will be assessed by a three-year Out-Of-Time rolling forecast where estimates are made based on all data available at time t which are then estimated on data from $t + 24$ to $t + 36$. For instance, using data from January 2000 up to December 2010, defaults from loans in January 2013 up to and including December 2013 are estimated. This way, a more complete view of performance when models are updated less regularly is obtained.

5.4.3 Model Evaluation

The estimated PDs for a given test set are then evaluated by comparing these to the observed defaults. Performance measures are calculated for each testing set. To provide a more balanced assessment of the PD models for the Out-Of-Time testing sets, these measures are weighted according to the number of observations in each year. This approach allows for a concise performance that considers the varying data volumes across different years. The Out-Of-Sample only contains only one testing set. Therefore, this weighting does not apply.

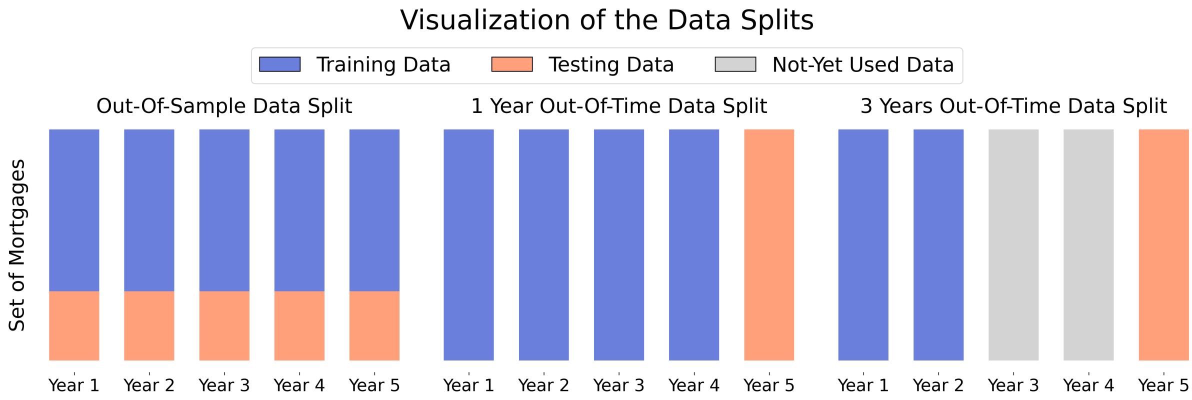


Figure 5: *Visualization of Train-Test Data Splits. In the first panel, 'Out-Of-Sample Data Split,' each year's data is divided into two parts, training (blue) and a fixed, smaller percentage for testing (orange). The 'One-Year Out-Of-Time Data Split' panel shows that data from the initial four years serve as the training set, with the final year reserved for testing. In the 'Three-Year Out-Of-Time Data Split' panel, only the first two years of data are utilized for training, while the fifth year is the test set; years three and four are not used yet, to resemble a model that has not been recalibrated on these data. These years are marked grey.*

6 Results

In this section, the results obtained from the estimation procedures in *PYTHON* will be discussed. The Out-Of-Sample results will be discussed in Subsection 6.1 and the Out-Of-Time results in Subsection 6.2. As discussed, the Out-Of-Time testing procedures have been conducted using both a one-year Out-Of-Time rolling forecast and a three-year Out-Of-Time rolling forecast. Both will have a dedicated subsection being Subsection 6.2.1 and Subsection 6.2.2 respectively.

As discussed previously, the performance measures are calculated for each year within the testing set separately. The values reported in this section are the weighted averages for the various testing sets. In particular for the Out-Of-Time results, this weighting approach is required to express the performance measure as a single index. Because some information is lost in this process, the three performance measures within each given year of the testing set are provided in Appendix C.4 in Table 12 through Table 23. To maintain consistency, within this appendix also the performance measures for the Out-Of-Sample model are reported for every year, but these are all calculated by filtering within the same set rather than being computed from a separate set.

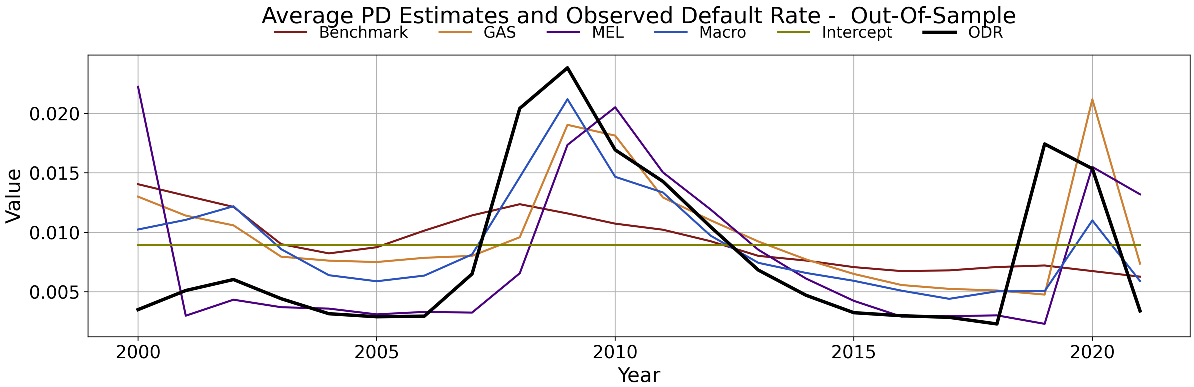
For the Out-Of-Sample results, the discussion will be based on a figure summarizing the performance metrics as well as on a table summarizing the regression output. For the Out-Of-Time results, the discussion shall be based on exclusively the figure summarizing the performance measures to avoid any unnecessary repetition. Moreover, the Out-Of-Time procedures have different regression outputs for every year rather than consisting of only one set of results, further complicating a structured discussion of the regression output directly.

As an extra point of reference, most results will also include an "intercept-only" model. In this model, for all mortgages j the PD is estimated by $p^j_{t,t+12} = \beta_0$. This does not mean that β_0 is necessarily equal across all periods as in the Out-Of-Time testing procedures this parameter is recalculated for each testing year.

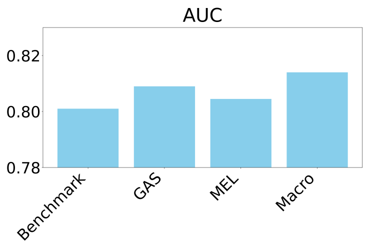
6.1 Out-Of-Sample Results

In Figure 6, the first summarizing figure is presented. Within Figure 6, there are four panels. In Figure 6a, the ODR over time is presented. In this panel, the average estimated default probability in each given calendar year can be compared to the realized default rate for all methods described in Section 3. In the lower row, the performance metrics themselves are presented. In Figure 6b the AUC from Equation (54) for the testing set is reported. In this panel, the intercept-only model is omitted since the estimated default probability in the testing set is not a function of the observed characteristics. Therefore, it can be argued that an intercept-only model does not have any discriminatory power or AUC value. In Figure 6c the average MSE over each individual mortgage is reported as in Equation (51). In Figure 6d the MSE of the models is computed by first computing the average of the estimated default probabilities in a given year and then comparing it to the ODR of that same year as is done in Equation (52).

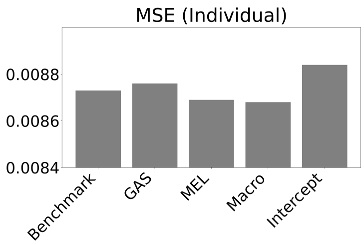
In Figure 12 in Appendix C.5, a figure showing the fluctuations of the estimated PDs over time is presented in a way alternative to Figure 6a. That is, in Figure 12 the values of $\hat{u}_t$ for the GAS, MEL and macroeconomic model are compared to the ODR.



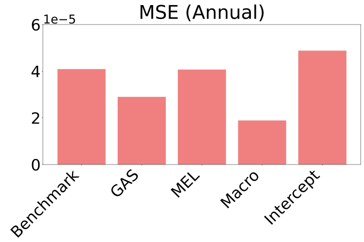
(a) *Observed 12-month Default Rate and Average Estimated 12-month PD Over Time*



(b) *AUC*



(c) *MSE (Individual)*



(d) *MSE (Annual)*

Figure 6: *Annual Observed and Estimated Default Rates and Histograms of Out-Of-Sample Predictions*

When considering the results in Figure 6, it can be remarked that the benchmark model in Figure 6a is more constant over time than the estimates from the other models and the ODR. As a consequence, the ODR is at some times larger than the benchmark estimate and at some times it is smaller. In terms of discriminatory power, in Figure 6b it can be seen that all other models attain greater AUC scores than the benchmark model. The individual MSEs in Figure 6c show that the benchmark model attains a greater MSE score and thus a smaller degree of calibration accuracy than the GAS model and the intercept-only model. When considering the MSE based on the annual averages of the estimated PDs in Figure 6d, the benchmark model is outperformed by both the GAS and the model with explicit macroeconomic characteristics. In turn, the MSE of the benchmark model is notably smaller than that of the intercept-only model. The difference between the MEL model is and the benchmark model is small, but the MSE of the benchmark model is slightly larger. If desired, this can be verified in Table 4, where it is reported that the MSE of the benchmark model is 0.04×10^{-5} units greater than that of the MEL model.

The estimates of the GAS model presented in Figure 6a show a pattern that closely resembles that of the observed average default rate. However, in particular in the period before 2008, the estimated default rates are larger than the realizations. The two spikes in default intensities in 2008 and 2019 are to some extent correctly estimated. However, the spikes in default estimates appear to be estimated one period later than the realizations. During 2019-2021, the GAS model stands out for being the only model surveyed estimating a maximum default rate that surpasses the largest observed rate in any of these years. In terms of the AUC in Figure 6b, the GAS model is outperformed only by the macroeconomic model. In Figure 6c, the result related to the benchmark model indicates that the benchmark model attains a greater score than the GAS and macroeconomic model when considering the individual MSE, but smaller than the MEL and the intercept-only model. The yearly average MSE is notably smaller than that of the benchmark model and the intercept-only model, but still greater than that of the macroeconomic model.

Naturally, the MEL model in Figure 6a has a similar shape as the ODR because of the large number of year dummies. The two spikes in default intensities are slightly underestimated. Moreover, the default rates appear to be lagging behind one year, as was also the case with the GAS model. This can be explained by Equation (40), where by definition the estimates $\hat{u}_t$ are based on data from up to and including period $t-12$. As another consequence of this, for the first year in the sample, there is no estimate for u_t, leading to the spike in estimated defaults in 2000. In Figure 6b, it can be seen that the discriminatory power of the GAS and the macroeconomic model is greater than that of the MEL model. The MSE based on the individual PDs is very similar but greater than that of the macroeconomic model in Figure 6c. Compared to the other three models, the MSE score is smaller. When considering the average annual MSE in Figure 6d, the MEL model attains a smaller MSE than the benchmark and the intercept-only model, but not than the GAS and the macroeconomic model.

The average probability of default estimates of the macroeconomic model in Figure 6a appear to resemble the ODR to a larger extent than the benchmark estimates. However, the model's predictions for 2009 and 2020 fall notably short of the actual default rates observed in those years. Like the GAS and the MEL model, the macroeconomic model reflected this increase in defaults only a year later and is in that sense lagging by one period. In Figure 6b, it can be seen that based on the AUC score, the macroeconomic model has a greater discriminatory power than the other models surveyed. In Figure 6c where the MSE scores based on individual mortgages are reported, the MSE of the macroeconomic model is the smallest. As previously noted, the difference with the MEL model is slim however. In Figure 6d, none of the models surveyed attains a smaller MSE score of the annual default rate estimates than the macroeconomic model.

Within the Out-Of-Sample results, it thus becomes clear from Figure 6 that the macroeconomic model attains the greatest degree of both discriminatory power and calibration accuracy model among the models included in this survey. This statement does not necessarily generalize to the macroeconomic model being the most reliable model for PIT PD estimation. In particular, in Section 6.2 it will be verified if the macroeconomic model still attains a greater AUC and smaller MSEs than the other models. Before proceeding with that, however, below the regression output of the models visualized in Figure 6 is included and will be discussed.

6.1.1 Regression Output

To provide some more detail regarding individual model performance, in Table 4 the regression outputs of the models are summarized. In this table, significance levels for risk drivers are determined by estimating standard errors using the variance-covariance matrix, Σ. The variance-covariance matrix is obtained by utilizing the fact that it equals the inverse of the negative Hessian matrix ($\Sigma = (-H)^{-1}$). This Hessian matrix H is the second derivative matrix of the log-likelihood function with respect to the parameters θ. For the logistic models, H is computed directly using its closed-form expression[10]. For the MEL and GAS models, such a closed-form expression of the Hessian matrix is not available so in these cases the Hessian matrix is estimated numerically using the output of PYTHON's *minimize* function. Several observations can be made from Table 4.

Concerning the significance and the directions of the estimates of the risk driver coefficients, in Table 4 the estimates for β are presented. The point estimates are mostly similar across all four models. All signs are in the same direction, except for "Loan Age", which is negative for the MEL model, but positive for the other models.

The macroeconomic risk drivers are all significant. For the GAS parameters, this is not the case. None of these are significant at the 10% level based on the t-statistics. What should be mentioned regarding this is that these standard errors are estimated by the Hessian matrix, which itself is also estimated. Therefore it is challenging to derive anything meaningful from this observation other than that this makes a slight

[10]The closed-form expression of the Hessian of a logistic regression model is given by $\frac{\partial^2 \ln L}{\partial\beta\partial\beta'} = -\sum_i \Lambda_i\left(1-\Lambda_i\right) x_i x_i'$ where $\Lambda_i = \Lambda\left(x'\beta\right) = \text{Prob}(Y=1 \mid x) = \frac{e^{x'\beta}}{1+e^{x'\beta}}$ (Seabold & Perktold, 2010)

implication of numerical instability. In Table 5, the year-dummies estimated in the MEL model are reported. There, most of the values are significant at the 0.1% level, with exceptions in 2008, 2012 and in 2015 through 2019.

	Benchmark Model Coefficients	**Macro E. model Coefficients**	**GAS Model Coefficients**	**MEL Model Coefficients**	**Intercept Only Model**
Intercept	-3.852***	-4.522***	-2.0710***	-4.785***	-4.708***
Original LTV	0.02247***	0.02248***	0.0122***	0.00209***	
Original UPB	0.00320***	0.00308***	0.0292***	0.00267***	
Original DTI Ratio	0.02439***	0.02294***	0.0212***	0.01944***	
Current Deferred UPB	-1.221×10^{-5} ***	-9.704×10^{-6} **	-1.088×10^{-5} ***	-1.461×10^{-5} **	
Credit Score	-0.00954***	-0.00969***	-0.0124***	-0.00949***	
Original Interest Rate	0.2825***	0.2474***	0.183***	0.5709***	
Channel Retail	-0.1327***	-0.1204***	-0.228***	-0.2013***	
Mortgage Ins. Pct	0.00263***	0.00326***	0.00696***	0.00138***	
Loan Term	0.00372***	0.00365***	0.0027***	0.00209***	
Loan Purpose No Cash-out	-0.4126***	-0.4034***	-0.235***	-0.2459***	
Loan Purpose Purchase	-0.5697***	-0.5366***	-0.222***	-0.4697***	
No. Borrowers	0.6438***	0.6369***	0.694***	0.5569***	
Loan Age	0.00292***	0.00280***	0.00248***	-0.00365***	
Channel Correspondent	-0.1068***	-0.08218**	-0.0522***	-0.2744***	
First Time Homebuyer	0.1171***	0.1155***	0.0036***	0.05678***	
CPI change (Current)		-2.398***			
GDP change (Lag 24 Month)		-0.6816**			
Unemployment Rate (Current)		0.08384***			
EPU (Lag 12 month)		0.000466***			
VIX (Current)		0.02423***			
u_0			-5.62×10^{-12}		
omega			0.0124		
B			-7.96×10^{-13}		
A			3.32×10^{-6}		
σ_u				0.9748	
Log Uncond. Likelihood (Eq. 38)				-121059	
LR Test Result				$p < 0.001$	
Individual PDs					
AUC (Test)	0.801	0.814	0.809	0.804	0.5
MSE (Test)	0.00873	0.00868	0.00876	0.00870	0.00884
AUC (Train)	0.802	0.814	0.809	0.838	0.5
MSE (Train)	0.00875	0.00869	0.00879	0.00867	0.00884
Average PDs					
MSE (Train)	4.09×10^{-5}	1.92×10^{-5}	2.88×10^{-5}	4.21×10^{-5}	4.88×10^{-5}
MSE (Test)	4.09×10^{-5}	1.89×10^{-5}	2.89×10^{-5}	4.05×10^{-5}	4.91×10^{-5}
Log-Likelihood (Test)	-121633	-119448	-119620	-120522	-136437
LR Test Result	×	$p < 0.001$	$p < 0.001$	$p < 0.001$	×

Table 4: *Summary of estimation results. Note: “***” indicates* $p < 0.01$*, “**” indicates* $p < 0.05$*, and “*” indicates* $p < 0.1$.

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Value	-1.855***	-1.307***	-1.014***	-0.817***	-0.933***	-0.997***	-1.112***	-0.436	0.741***	1.060***	0.850***

Year	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Value	0.795	0.662***	0.452***	0.215	-0.040	0.012	0.006	-0.247	1.832***	1.823***

Table 5: *MEL Estimates of* u_t *for 2001-2021. Note: “****” indicates* $p < 0.01$*, “**” indicates* $p < 0.05$*, and “*” indicates* $p < 0.1$.

6.1.2 Likelihood Ratio Tests Results

In the row denoted by Log Uncond. Likelihood (Eq. 38) in Table 4, the log-likelihood that is unconditional with respect to u_t is reported for the MEL model. The hypotheses of the test conducted here are $H_0 : \sigma_u = 0$ and $H_1 : \sigma_u > 0$. Since the benchmark model is nested in the MEL model given that the models are equivalent under H_0, a likelihood ratio test can be conducted where the test statistic is given by Equation (55). The test result is reported in Table 6 below.

In the lower rows of Table 4, the likelihood of the testing sets of the models is reported. In the row below, the following test is conducted: $H_0 : u_t = 0$ for all t or $H_a : u_t \neq 0$ for at least one period t. Since the benchmark model is nested in the macroeconomic model, in the GAS model and in the MEL model, a likelihood ratio test can again be conducted with the likelihood ratio test statistic from Equation (55). Given the large differences in likelihood and thereby clear rejections of H_0, this shall not be discussed in further detail. In Table 6 the test statistics, degrees of freedom, critical values at the 0.1% level and conclusions are reported.

Model	**Test Statistic**	**df**	**Critical value at 0.1% Level**	**Conclusion**
MEL (unconditional)	1148	1	10.83	Reject $H_0 : \sigma = 0$
GAS	4290	3	16.27	Reject $H_0 : u_t = 0$
Macroeconomic	4026	5	20.51	Reject $H_0 : u_t = 0$
MEL (conditional)	2222	22	48.27	Reject $H_0 : u_t = 0$

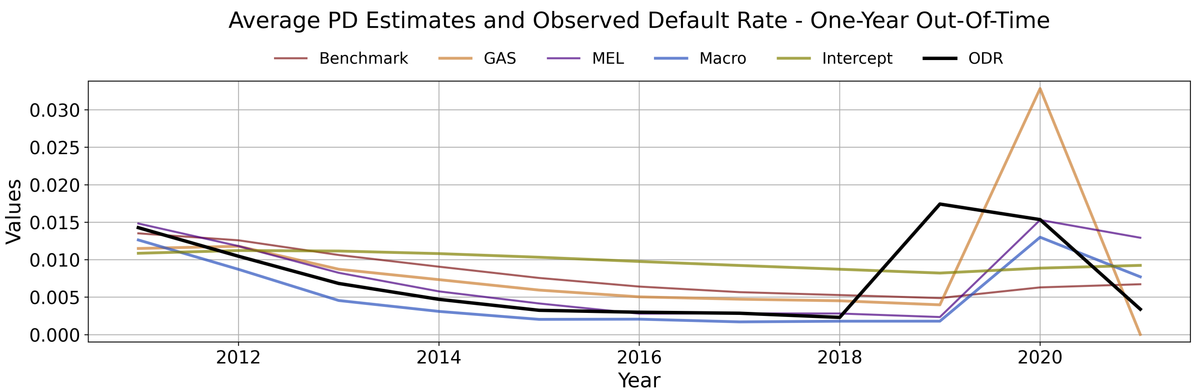
Table 6: *Summary results of likelihood ratio tests.*

It should be noted that this test is performed across all periods. Therefore, it is not designed to reflect whether in a certain subset of t, it would have been justified to exclude u_t. The results from this test should thus be interpreted with appropriate care.

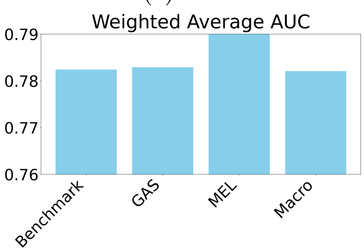
Based on the results in Figure 6, Table 4, Table 5 and Table 6, it is clear that the effect of the state-of-the-economy is successfully integrated into each of the models aiming to do so. This can be derived from the fact that the models that do include u_t, MEL, the macroeconomic model and the GAS model display appropriately timed spikes in default intensities in the correct direction. However, all three of these models exhibit a one-year lag before these increases in the default rate are reflected. Moreover, the results from the likelihood ratio tests are all in favor of this conclusion.

6.2 Out-Of-Time Results

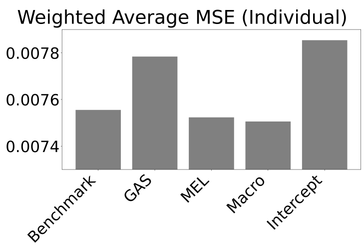
6.2.1 One-Year Out-Of-Time Results



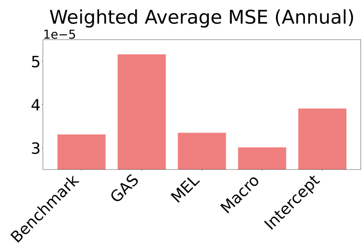
(a) *Observed 12-month Default Rate and Average Estimated 12-month PD Over Time*



(b) *Weighted Average AUC*



(c) *Weighted Average MSE (Individual)*



(d) *Weighted MSE (Annual)*

Figure 7: *Annual Observed and Estimated Default Rates and Histograms of 1 Year Out-Of-Time Predictions*

In Figure 7, the results for the one-year Out-Of-Time test are reported. As in the OOS test, in Figure 7a, the estimated PDs are plotted together with the ODR. In Figure 13 in Appendix C.5, a similar figure is shown where $\hat{u}_t$ of the GAS, MEL, and macroeconomic model compared to the ODR. In Figure 7b, the weighted averages of the AUC scores of the models are reported. In Figure 7c and 7d the same is done for the MSE of the individual mortgages and the yearly averages respectively.

When considering the benchmark model, it can be seen in Figure 7a that the default rate is relatively constant over time. The spike around the COVID-19 crisis is not reflected in these estimates. In Figure 7b, the AUC score of this model is nearly identical to that of the GAS model and the macroeconomic model, but notably smaller than that of the MEL model. When comparing the MSE based on the individual mortgages of the benchmark in Figure 7c, it can be seen that this score is smaller for the benchmark model than for the GAS model and the intercept-only model. However, compared to the MSEs of the MEL and macroeconomic model it is slightly greater. In Figure 7d, it is shown that after the macroeconomic model, the benchmark model attains the smallest MSE based on the yearly average estimated PDs. That said, the difference with the MEL model is small.

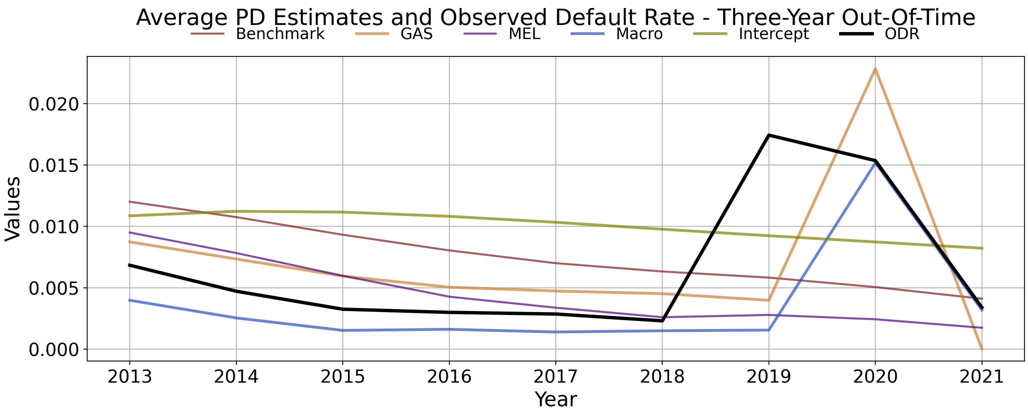
What stands out for the GAS estimates in Figure 7a is that although the GAS model is able to reflect the increased default rate in 2020, the estimated default rate is much larger than the realization. Conversely, in 2021, the estimated default rate of the GAS model is much smaller than its realization. Both of these years contribute to the GAS model having the largest MSE in Figure 7c after the intercept-only model and the largest over all models in Figure 7d. The AUC of the GAS model in Figure 7b is nearly identical to that of the benchmark model and of the macroeconomic model and smaller than that of the MEL model.

Similar to the cases before, the MEL model exhibits a one-year lag in estimating the spike in default intensities, but the magnitude of this spike is estimated more accurately than by the GAS model. As noted before, the AUC of the MEL model is the greatest among the models surveyed in this case. In Figure 7c and Figure 7d, it can be seen that based on the two MSE scores, the MEL model is in this sample only outperformed by the macroeconomic model.

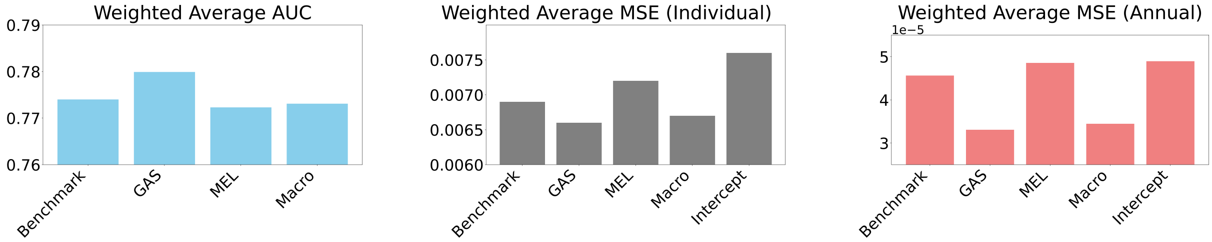
The macroeconomic model exhibits a similar pattern as the MEL model in Figure 7a, but the estimate in 2021 is more similar to the ODR than that of the MEL model. This is reflected in Figure 7c and Figure 7d, where the macroeconomic model attains the smallest score for both variations of the MSE and therefore the degree largest calibration accuracy. In Figure 7b it can be seen that this is not the case for the discriminatory power, as the AUC of the MEL model is larger.

To conclude, in the one-year Out-Of-Time test, the macroeconomic model again attains the smallest MSE scores in Figure 7c and 7d. However, in Figure 7b it can be seen that the MEL model has the largest discriminatory power as its AUC is largest.

6.2.2 Three-Year Out-Of-Time Results



(a) *Values of Observed 12-month Default Rate and Average Estimated 12-month PD Over Time*



(b) *Weighted Average AUC*

(c) *Weighted Average MSE (Individual)*

(d) *Weighted MSE (Annual)*

Figure 8: *Annual Observed and Estimated Default Rates and Histograms of three-year Out-Of-Time Predictions*

In Figure 8, the results for the three-year Out-Of-Time testing procedure are reported using the same four-panel structure as before. In Figure 8a, the annual average of the estimated default rates for the three-year Out-Of-Time test are reported. Similar to before, in Figure 14 in Appendix C.5, a figure is shown for $\hat{u}_t$ compared to the ODR, there, the same pattern is visible. IN Fiugre Figure 8b, the weighted averages of the AUC scores are reported. The MSEs of both the individual PDs and the yearly average PDs are presented in Figure 8c and Figure 8d respectively.

For the benchmark model, as shown in Figure 8a, the estimated default rate over time is stable. In Figure 8b, its AUC score closely matches that of the MEL and macroeconomic models, although out of these three, the AUC of the benchmark model is the largest by a small amount. Moreover, the AUC is notably smaller than that of the GAS model. When examining the MSE for individual mortgages in Figure 8c, the benchmark model attains a score smaller than those of the MEL and intercept-only models, but greater than those of the GAS and macroeconomic models. In terms of yearly average estimated PDs, Figure 8d shows that the benchmark model achieves a comparatively large MSE when compared to that of the GAS and the macroeconomic model.

Regarding the GAS model, Figure 8a once more demonstrates its ability to capture the increased default rate in 2020, although this default rate is still overestimated. However, the extent of this overestimation is notably smaller than in Figure 7a. In Figure 8b it is shown that in this case the AUC is maximized by the GAS model. The more accurate estimate of the default spike around the COVID-19 crisis may contribute to the GAS model having the smallest MSEs as shown in Figure 8c and Figure 8d. However, for both MSEs, the difference between the GAS and macroeconomic model is small in this case. In Table 22 and Table 23 in Appendix C.3.1, the corresponding values are reported and it can be verified that the GAS model indeed attains two slightly smaller MSE scores.

In Figure 8a, the MEL model exhibits behavior that is different from that in Figure 6a and Figure 7a. The important difference is that in Figure 8a, the spike in defaults around 2019 is not reflected. This is again a result of Equation (40). In this setup $\hat{u}_t$ is estimated using data up until period $t-36$ rather than $t - 12$. In Figure 8b the AUC score of the MEL model is the smallest among the models. Figures 8c and 8d indicate that the MEL model attains the second largest MSE scores, being smaller than only that of the intercept-only model.

Finally, it can be seen in Figure 8a that the macroeconomic model can reflect the spike in defaults in 2019 to a similar extent as in the Out-Of-Sample and in the one-year Out-Of-Time testing procedure. The AUC in Figure 8b is, as noted, similar to that of the benchmark and MEL model and smaller than that of the GAS model. In Figures 8c and 8d the macroeconomic model shows that the MSEs of the macroeconomic model are slightly greater for the macroeconomic model than for the GAS model and notably smaller than those of the other alternatives.

To sum up, in the three-year Out-Of-Time testing procedure, based on all three performance measures considered the GAS model is preferred. Together with the macroeconomic model, the differences with the other models considered are relatively large. Among these two models, however, the difference in terms of MSE is small compared to the size of the difference with the other models.

7 Discussion

In this thesis, different methods for estimating the 12-month PIT PD were explored and evaluated using Freddie Mac mortgage data from 2000 to 2022. In Section 2 it was noted that the definition of a "Point-In-Time" probability of default has been the center of debate for some decades. The definition that is maintained in this thesis is provided in Equation (4): the PIT is a probability of default conditional on the current state of the economy.

The main objective of the thesis is to determine which of the models surveyed should be used to estimate this PIT PD most reliably in the context of US mortgages. The models surveyed are all logistic models where the state-of-the-economy is estimated differently. The methods used to this extent are the GAS recursion by Creal et al. (2013), the MEL model and the incorporation of macroeconomic data as separate risk drivers. Moreover, a benchmark model was included to verify whether any of these estimates for the state-of-the-economy yield any improvement in model performance. This analysis was designed to address the following overarching research question:

Research Question: *Which of the following methods for estimating the state-of-the-economy yields the most reliable Point-In-Time PD model for U.S. mortgages: using a linear combination of macroeconomic indicators, implementing a GAS recursion, using a Mixed Effects Logistic model, or simplifying the model by not incorporating any state-of-the-economy estimate?*

Purely based on statistical measures, no definitive answer to this question can be provided based on the results in this thesis. However, in the discussion below this question will be answered to some extent.

First of all, based on the MSE scores in the results, it can be concluded that using the macroeconomic model or the GAS model leads to higher calibration accuracy in all cases assessed as compared to the MEL and benchmark model. Therefore, when the practitioner aims to minimize estimation error, it can be concluded, based on the results in this thesis, that the GAS and the macroeconomic model are better suited for modeling the PD of US mortgages. Still, it is unclear which of these two models should be used under this objective.

This ambiguity stems largely from the fact that the GAS model performance varies across different tests. Specifically, in the three-year Out-Of-Time test, the GAS model outperforms all other models based on the two MSEs. This preference is absent in the one-year Out-Of-Time and Out-Of-Sample tests. In these cases, the macroeconomic model demonstrates superior performance based on the two MSEs.

Similarly, the results in this thesis are not unambiguous when determining which of the models should be used when the target is to maximize discriminatory power. In all three tests considered, the discriminatory power is maximized by another model.

Having discussed the performance measures of the various models that do include an estimate for the state-of-the-economy, the second consideration in the research question arises: is it beneficial to incorporate an estimate for the state-of-the-economy in this PIT PD modeling context? Considering that the benchmark where u_t is excluded consistently obtains smaller AUC scores and larger MSE scores throughout most of the results, it is suggested that the answer to this question is yes. When the target is to optimize these measures, all models yield an improvement over the benchmark case, so incorporating any of the surveyed estimation procedures for the state-of-the-economy may be worthwhile. Still, it deserves to be stressed that this need not necessarily be generalizable to other data. Moreover, from the likelihood ratio tests in the Out-Of-Sample test, it became clear that the models including u_t are preferred in at least Out-Of-Sample analysis. Additionally, based on the LR-test conducted on σ in the MEL model, it is suggested that some structural noise is present, adding some extra justification for the modeling of this term.

Contrary to the statistical performance measures discussed earlier, the differences among the models concerning practicality and explainability are quite clear.

First of all, the GAS model as defined in this thesis is highly sensitive to the choice of starting values in the sense that estimation fails due to numerical issues in most cases. This numerical issue can clearly be seen in Table 4, where none of the GAS parameters is significant. This might happen either because these parameters are truly insignificant, or because the estimated standard errors are not reliable due to these being estimated based on a Hessian matrix, which itself is also estimated. Regardless of the underlying cause, the fact that this issue arises at all highlights the need for careful application of the GAS model. Another important obstacle is the definition of the gradient. In this thesis, this has been defined somewhat naively by Equation (48). When this gradient is not included, the results are even less stable and the numerical issues worsen. However, it may be used as an objection to the validity of this model that this gradient has been altered in this way. Nevertheless, the approximation does allow for more numerical stability as is its primary aim and the results seem to resemble the data at least to some extent.

For the model including macroeconomic characteristics, there are no noteworthy limitations within the estimation procedure itself. However, the limitations are in this case not in the estimation process, but within the macroeconomic data used to model the state-of-the-economy. As noted in Section 3.2.3, in Koopman et al. (2011) it was shown that regardless of the number and the quality of the macroeconomic risk drivers, a significant part of the variation in the default rate remains unexplained. More recently, during COVID-19 in the EU, macroeconomic values should have resulted in a strongly increased PD, yet due to government subsidies, the default rates did not increase substantially. This shortcoming is also reflected by the results in Section 6 and should be taken into consideration when determining which of the models can estimate the PIT PD most reliably.

A limitation of the MEL model can clearly be seen in the three-year Out-Of-Time results. There, the MEL model fails to register the increase in observed defaults. The reason why this occurs is that the MEL model uses an estimate of the state of the economy which is already three years old by the time the three-year Out-Of-Time test estimates are constructed. Consequently, if macroeconomic conditions change in the meantime, this is not reflected. This strong dependence on calibration frequency is arguably the largest disadvantage of this model.

In the remainder of this discussion, the results regarding the conclusion based on the performance measures and the findings regarding practicality will be combined to provide an answer to the research question, which is to determine which of the surveyed models can provide the most reliable estimate of the PIT PD. The benchmark model is left out of consideration because it is almost consistently outperformed by the other models and the likelihood ratio tests support the inclusion of u_t. Moreover, the MEL model is preferred only when considering the AUC in the one-year Out-Of-Time case, and in addition, this model cannot always reflect underlying changes in the economy as can be seen in the three-year Out-Of-Time case. Consequently, the research question is reduced to whether the GAS model or the macroeconomic model is better suited for estimating the PIT PDs of US mortgages reliably.

The answer to this question is as follows: Given the practical limitations of the GAS model, there is no good reason to say with certainty that the GAS methodology should be used over the macroeconomic model given that sufficient and most importantly, relevant macroeconomic data is present. Conversely, purely based on performance, the GAS model may provide a viable alternative to estimate the PD when macroeconomic data are not available or not reliable. At any rate, the GAS model is not consistently worse-performing than the macroeconomic model based on the performance measures considered in this thesis and no definitive preference can be derived.

So based on the results in this thesis, either the GAS model or the model including macroeconomic characteristics models could be used to obtain reliable PIT PD estimates and it is up to the practitioner to decide which model is best suited for their specific question. Yet, it is important to consider the external validity of these findings; the effectiveness of each model may vary depending on different datasets and economic environments, so careful evaluation is necessary when applying these models to other contexts.

7.1 Limitations and Future Research

It is essential to consider the context and conditions under which this research was conducted when applying its conclusions more broadly. For that reason, a few of the research limitations will be listed in this subsection.

The first limitation of this thesis is that in the GAS model, some computational difficulties have not been resolved in full, potentially leading to the results yielded by this model being less accurate. This is in particular true for the gradient function, which has been somewhat simplified in order to proceed. A similar point can be made regarding the standard errors of the GAS model. Although these are less likely to degrade the overall quality of the model, they do introduce a degree of unreliability. As a result, the conclusions drawn from this thesis regarding this model should be approached with extra caution. For future research, it would be interesting to verify whether one would arrive at similar conclusions as those presented in this thesis if these computational issues are resolved in full.

Secondly, the way in which the probability of default itself has been estimated in this thesis is somewhat simplified. Throughout this thesis, it is assumed that only two loan classes exist, being loans that are in default and loans that are not in default. However, often a more nuanced approach is taken, where loans are categorized into various “buckets” indicating different levels of default likelihood. Using such approaches, loans can, over different periods, move not only from a state of non-default to default but also between different buckets. This approach is discussed thoroughly in textbooks on credit risk management such as Duffie & Singleton (2012) and also in some of the papers referenced in this thesis, such as Gavalas & Syriopoulos (2014) and Kiefer & Larson (2014). Given the increased complexity of the bucketed approach, this thesis intentionally adopts a simplified model focusing on two loan classes to provide a clear and manageable framework for analyzing different PIT PD models. For future research, it could be examined how the models in this thesis perform under such a setting.

Thirdly, the fact that the PDs are constructed yearly might have affected the estimation results. It is reasonable to assume that in some instances, such as the financial crisis or the COVID-19 crisis, PDs vary strongly not only from year to year, but also from quarter to quarter or even from month to month. Such a change in updating frequency could particularly benefit the MEL model, since the estimates of the state-of-the-economy are based on observations from the previous period. If these periods are shorter, the data that are used to construct this estimate are less outdated, potentially enhancing its performance. Investigating how more frequent updates of the estimates would affect model performance thus not only forms a limitation, but also presents an interesting research opportunity.

Moreover, it would be interesting to see how the results from the models assessed in this thesis generalize to other geographical locations. For example, unemployment during the COVID crisis rose dramatically in the US. The extent of this in the EU was relatively mild. As such, one might expect that the findings from this period could vary significantly in an EU context. Especially since contrary to in the US, default rates increased less sharply in Europe due to government support. Similarly, the applicability of this research might be limited when considering different types of mortgages, geographic regions, or periods.

Taking all these points into account, this thesis marks a step forward in our understanding of estimating default probabilities, while also paving the way for further studies on mortgage risk in varying economic conditions. There is much more to be learned in this field, and it will be fascinating to see how this area of research evolves in the future.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. Petrov & F. Csaki (Eds.), *2nd international symposium on information theory* (pp. 267–281). Budapest, Hungary: Akadémiai Kiadó. (is a chapter of 2990687 2nd international symposium on information theory, Tsahkadsor, Armenia, USSR, September 2-8, 1971)

Albuquerque, B., & Varadi, A. (2023). Role of government policies in smoothing borrowers' spending during stress: Evidence from uk mortgage moratoria. *SSRN Electronic Journal*. (Available at SSRN: https://ssrn.com/abstract=4359121 or http://dx.doi.org/10.2139/ssrn.4359121) doi: 10.2139/ssrn.4359121

Bain, L. J., & Engelhardt, M. (1992). *Introduction to probability and mathematical statistics* (2nd ed.). Duxbury/Thomson Learning.

Baker, S. R., Bloom, N., & Davis, S. J. (2016, 07). Measuring Economic Policy Uncertainty*. *The Quarterly Journal of Economics*, *131*(4), 1593-1636. Retrieved from `https://doi.org/10.1093/qje/qjw024` doi: 10.1093/qje/qjw024

BCBS. (2005). *Studies on the validation of internal rating systems (revised)* (Working Paper No. 19). Bank for International Settlements. Retrieved from `https://www.bis.org/publ/bcbs_wp19.htm` (Status: Current)

BCBS. (2006). *International convergence of capital measurement and capital standards: A revised framework, comprehensive version.* Bank for International Settlements. (This document is a compilation of the June 2004 Basel II Framework, the elements of the 1988 Accord that were not revised during the Basel II process, the 1996 Amendment to the Capital Accord to Incorporate Market Risks, and the 2005 paper on the Application of Basel II to Trading Activities and the Treatment of Double Default Effects. No new elements have been introduced in this compilation.)

BCBS. (2010, December). Basel iii: A global regulatory framework for more resilient banks and banking systems.
(Revised in June 2011)

BCBS. (2019, Dec 15). *Cre36 - irb approach: Minimum requirements to use irb approach.* CRE - Calculation of RWA for Credit Risk. (This chapter sets out the minimum requirements for banks to use the internal ratings-based approach, including requirements for initial adoption and for ongoing use.)

BCBS. (2023). *History of the basel banking accords.* Retrieved from `https://www.bis.org/bcbs/history.htm`

Biljanovska, N., Chen, S., Gelos, G., Igan, D., Martinez Peria, M. S., Nier, E., & Valencia, F. (2023). *Macroprudential policy effects: Evidence and open questions* (Working Paper No. DP/2023/002). International Monetary Fund.

Blasques, F., Creal, D. D., Janus, P., Koopman, S. J., Lucas, A., Scharth, M., & Schwaab, B. (2014). Generalized autoregressive score models for time-varying parameters: new models and applications. In *Workshop isf 2014 rotterdam.*

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications.* Cambridge University Press. doi: 10.1017/CBO9780511811241

Campbell, J. Y., & Cocco, J. F. (2015). A model of mortgage default. *The Journal of Finance*, *70*(4), 1495-1554. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12252` doi: https://doi.org/10.1111/jofi.12252

Cao, R., Vilar, J. M., & Devia, A. (2009). Modelling consumer credit risk via survival analysis. *SORT: statistics and operations research transactions*, *33*(1), 0003–30.

Carlehed, M., & Petrov, A. (2012). A methodology for point-in-time-through-the-cycle probability of default decomposition in risk classification systems. *Journal of Risk Model Validation*, *6*(3), 3–25. Retrieved from `https://tilburguniversity.on.worldcat.org/oclc/8096829908` doi: 10.21314/JRMV.2012.091

Castro, V. (2013). Macroeconomic determinants of the credit risk in the banking system: The case of the gipsi. *Economic Modelling*, *31*, 672-683. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0264999313000308` doi: https://doi.org/10.1016/j.econmod.2013.01.027

Chawla, G., Forest Jr, L. R., & Aguais, S. D. (2017). Convexity and correlation effects in expected credit loss calculations for ifrs9/cecl and stress testing. *Journal of Risk Management in Financial Institutions*, *10*(1), 99-110. Retrieved from `https://tilburguniversity.on.worldcat.org/oclc/6941513747`

Chi, Q., & Li, W. (2017). Economic policy uncertainty, credit risks and banks' lending decisions: Evidence from chinese commercial banks. *China Journal of Accounting Research*, *10*(1), 33-50. Retrieved from `https://www.sciencedirect.com/science/article/pii/S1755309116300466` doi: https://doi.org/10.1016/j.cjar.2016.12.001

Cournapeau, D., Pedregosa, F., Varoquaux, G., Gramfort, A., & Michel, V. (2023). *scikit-learn: Machine learning in Python.* Retrieved from `https://scikit-learn.org/stable/` (Accessed: 2023-11-19)

Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, *28*(5), 777-795. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.1279` doi: https://doi.org/10.1002/jae.1279

Crotty, J. (2009, 07). Structural causes of the global financial crisis: a critical assessment of the 'new financial architecture'. *Cambridge Journal of Economics*, *33*(4), 563-580. Retrieved from `https://doi.org/10.1093/cje/bep023` doi: 10.1093/cje/bep023

Demidenko, E. (2013, 11). Mixed models. theory and applications with r. 2nd ed. doi: 10.1002/9781118651537

Duffie, D., & Singleton, K. J. (2012). *Credit risk: Pricing, measurement, and management.* Princeton University Press. Retrieved from `file:///C:/Users/ssentjens/Downloads/10.1515.9781400829170%20(1).pdf` (Retrieved May 8, 2023)

Dutch Ministry of Finance. (2022). *Besluit noodmaatregelen coronacrisis.* Besluit van de Staatssecretaris van Financiën. Retrieved from `https://wetten.overheid.nl/BWBR0046255/2022-04-01` (Geldend van 01-04-2022 t/m 19-09-2022. Besluit is een actualisatie van het besluit van 16 december 2021, nr. 2021-258581 (Stcrt. 2021, 50389), laatst gewijzigd bij besluit van 19 januari 2022, nr. 2022-12961 (Stcrt. 2022-1588).)

EBA. (2017, 11 20). *Report on irb modelling practices: Impact assessment for the gls on pd, lgd and the treatment of defaulted exposures based on the irb survey results* (EBA Report). European Banking Authority. (Available at EBA website)

EBA. (2020, April). *Final report on guidelines on legislative and non-legislative payment moratoria* (Guidelines No. EBA/GL/2020/02). European Banking Authority. (Guidelines on legislative and non-legislative moratoria on loan repayments applied in the light of the COVID-19 crisis)

EBA. (2022a, November 4). *2023 eu-wide stress test – methodological note* (Tech. Rep.). European Banking Authority.

EBA. (2022b). Article 154: Risk-weighted exposure amounts for retail exposures. *Capital Requirements Regulation (CRR)*, *PART THREE*(TITLE II, CHAPTER 3, Section 2, Sub-Section 2, Article 154). Retrieved from `URLoftheDocument(ifavailable)` (Provide additional information or context here if necessary)

Eder, B. (2021, December 9). Revisiting the dualism of point-in-time and through-the-cycle credit risk measures. *SSRN*. Retrieved from https://ssrn.com/abstract=3981340 doi: 10.2139/ssrn.3981340

Figlewski, S., Frydman, H., & Liang, W. (2006, 09). Modeling the effect of macroeconomic factors on corporate default and credit rating transitions. *International Review of Economics & Finance*, *21*. doi: 10.2139/ssrn.934438

Foote, C., Gerardi, K., Goette, L., & Willen, P. (2010). Reducing foreclosures: No easy answers. *NBER Macroeconomics Annual*, *24*, 89-138. Retrieved from https://doi.org/10.1086/648289 doi: 10.1086/648289

Foote, C., Gerardi, K., & Willen, P. S. (2008). Negative equity and foreclosure: Theory and evidence. *Journal of Urban Economics*, *64*(2), 234-245. Retrieved from https://www.sciencedirect.com/science/article/pii/S0094119008000673 doi: https://doi.org/10.1016/j.jue.2008.07.006

Freddie Mac. (2023, October). Single-family loan-level dataset general user guide [Computer software manual]. Retrieved from http://www.freddiemac.com/research/datasets/sf_loanlevel_dataset.html (Accessed: 2023-11-06)

Gavalas, D., & Syriopoulos, T. (2014). Bank credit risk management and rating migration analysis on the business cycle. *International Journal of Financial Studies*, *2*(1), 122–143. Retrieved from https://www.mdpi.com/2227-7072/2/1/122 doi: 10.3390/ijfs2010122

Gerardi, K., Shapiro, A. H., & Willen, P. S. (2007). *Subprime outcomes: risky mortgages, homeownership experiences, and foreclosures* (Working Papers No. 07-15). Federal Reserve Bank of Boston. Retrieved from https://ideas.repec.org/p/fip/fedbwp/07-15.html

Guo, L., & Bruneau, C. (2014). Macroeconomic variables and default risk: An application of the favar model. *Revue d'économie politique*, *124*(5), 817-857. Retrieved from https://www.cairn.info/revue-d-economie-politique-2014-5-page-817.htm doi: 10.3917/redp.245.0817

IASB, I. A. S. B. (2021). *Ifrs 9 financial instruments* (2021edition ed.).

Kiefer, N., & Larson, C. (2014, 10). Testing simple markov structures for credit rating transitions.

Koopman, S. J., & Lucas, A. (2005a). Business and default cycles for credit risk. *Journal of Applied Econometrics*, *20*(2), 311–323. Retrieved 2023-09-20, from http://www.jstor.org/stable/25146357

Koopman, S. J., & Lucas, A. (2005b). Business and default cycles for credit risk. *Journal of Applied Econometrics*, *20*(2), 311–323. Retrieved 2023-11-16, from http://www.jstor.org/stable/25146357

Koopman, S. J., Lucas, A., & Schwaab, B. (2011). Modeling frailty-correlated defaults using many macroeconomic covariates. *Journal of Econometrics*, *162*(2), 312-325. Retrieved from https://www.sciencedirect.com/science/article/pii/S0304407611000303 doi: https://doi.org/10.1016/j.jeconom.2011.02.003

Mageto, D. K., Mwalili, S. M., & Waititu, A. G. (2015). Modelling of credit risk: Random forests versus cox proportional hazard regression. *American Journal of Theoretical and Applied Statistics*, *4*(4), 247–253.

Mcculloch, C. E., & Neuhaus, J. M. (2013). Generalized linear mixed models. In *Encyclopedia of environmetrics.* John Wiley & Sons, Ltd. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470057339.vag009.pub2 doi: https://doi.org/10.1002/9780470057339.vag009.pub2

McKinney, W. (2023). *pandas: powerful python data analysis toolkit.* Retrieved from https://pandas.pydata.org/pandas-docs/stable/ (Accessed: 2023-11-19)

Medema, L., Koning, R. H., & Lensink, R. (2009). A practical approach to validating a pd model. *Journal of Banking & Finance*, *33*(4), 701-708. Retrieved from https://www.sciencedirect.com/science/article/pii/S0378426608002781 doi: https://doi.org/10.1016/j.jbankfin.2008.11.007

Miu, P., Ozdemir, B., & Ozdemir, M. (2005, 01). Practical and theoretical challenges in validating basel parameters: key learnings from the experience of a canadian bank. *Journal of Credit Risk*, *1*, 89-136. doi: 10.21314/JCR.2005.026

Oliphant, T. (2023). *Numpy.* Retrieved from `https://numpy.org/doc/` (Accessed: 2023-11-19)

Pennington-Cross, A., & Ho, G. (2010). The termination of subprime hybrid and fixed-rate mortgages. *Real Estate Economics*, *38*(3), 399-426. Retrieved from `https://EconPapers.repec.org/RePEc:bla:reesec:v:38:y:2010:i:3:p:399-426`

Roncalli, T. (2020). *Handbook of financial risk management.* doi: 10.1201/9781315144597

Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. In *9th python in science conference.*

Simons, D., & Rolwes, F. (2009, Sep). Macroeconomic default modeling and stress testing.

Tasche, D. (2006). *Validation of internal rating systems and pd estimates.*

Treacy, W., & Carey, M. (1998, 01). Credit risk rating systems at large us banks. *Journal of Banking& Finance*, *24*, 167-201. doi: 10.1016/S0378-4266(99)00056-4

Xia, Y., He, L., Li, Y., Fu, Y., & Xu, Y. (2021). A dynamic credit scoring model based on survival gradient boosting decision tree approach. *Technological and Economic Development of Economy*, *27*(1), 96–119. Retrieved from `https://doi.org/10.3846/tede.2020.13997` doi: 10.3846/tede.2020.13997

Yang, W. Y., Cao, W., Kim, J., Park, K. W., Park, H.-H., Joung, J., … Im, T. (2020). *Applied numerical methods using matlab, 2nd edition.* John Wiley & Sons, Inc.

Zhu, C., Byrd, R., Lu, P., & Nocedal, J. (1997, December). Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software*, *23*(4), 550–560. doi: 10.1145/279232.279236

Appendices

A Gauss-Hermite Quadrature

This discussion of Gauss-Hermite Quadrature shall be based on Section 5.9.2 of Yang et al. (2020), although the contents discussed there are universal.

Gauss-Hermite Quadrature is a method that can be used to numerically evaluate an integral by evaluating that function at predetermined nodes and corresponding weights as in

$$\int_{-\infty}^{+\infty} e^{-x^2} f(x)dx \approx \sum_{i=1}^{n} w_i f(x_i). \tag{56}$$

The $n-th$ grid point x_i can be obtained as the zeroes of the n-point Hermite polynomial given by

$$H_n(x) = (-1)^n e^{-x^2} \frac{d^n}{dx^n} e^{-x^2}. \tag{57}$$

And the associated weights are then given by

$$w_i = \frac{2^{n-1} n! \sqrt{\pi}}{n^2 \left[H_{n-1}(x_i)\right]^2}. \tag{58}$$

B Derivations and Formulae

B.1 Coefficient of Correlation from Section 2

In Equation (1) in Section 2, it was mentioned that the coefficient of correlation R is a function of the PD, but the function was omitted there for brevity. For completeness' sake, this function (as it is configured for consumer mortgages in (EBA, 2022b)) is provided below

$$R = 0.03 \cdot \frac{1 - e^{-35 \cdot PD}}{1 - e^{-35}} + 0.16 \cdot \left(1 - \frac{1 - e^{-35 \cdot PD}}{1 - e^{-35}}\right). \tag{59}$$

Here, R is the coefficient of correlation and PD is the probability of default.

B.2 Derivation of Log-Likelihood of the Benchmark PD Model

In this appendix, the derivation of the log-likelihood of the Benchmark model given in Equation (30) in Section 4.1 is provided. Broken down step-by-step this derivation is given by

$$\begin{aligned}
\log L(\beta \mid X_{\mathcal{T}}, D_{\mathcal{T}}) &= \sum_{t\in\mathcal{T}} \sum_{j\in L_t} \left[D^j_{t+12} \log \mathbb{P}(D^j_{t+12} = 1 \mid u_t, X^j_t, D^j_{t+12} = 0) + \left(1 - D^j_{t+12}\right) \log\left(1 - \mathbb{P}(D^j_{t+12} = 1 \mid u_t, X^j_t, D^j_{t+12} = 0)\right)\right] \\
&= \sum_{t\in\mathcal{T}} \sum_{j\in L_t} \left[D^j_{t+12} \log\left(\frac{1}{1 + e^{-\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt}}}\right) + \left(1 - D^j_{t+12}\right) \log\left[\frac{e^{-\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt}}}{1 + e^{-\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt}}}\right)\right] \\
&= \sum_{t\in\mathcal{T}} \sum_{j\in L_t} \left[D^j_{t+12} \left\{\log\left(\frac{1}{1 + e^{-(\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt})}}\right) - \log\left(\frac{e^{-(\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt})}}{1 + e^{-(\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt})}}\right)\right\} + \log\left(\frac{e^{-(\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt})}}{1 + e^{-(\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt})}}\right)\right] \\
&= \sum_{t\in\mathcal{T}} \sum_{j\in L_t} \left[D^j_{t+12} \left\{\log\left(e^{(\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt})}\right)\right\} + \log\left(\frac{e^{-(\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt})}}{1 + e^{-(\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt})}} * \frac{e^{(\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt})}}{e^{(\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt})}}\right)\right] \\
&= \sum_{t\in\mathcal{T}} \sum_{j\in L_t} \left[D^j_{t+12} (\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt}) + \log\left(\frac{1}{1 + e^{(\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt})}}\right)\right] \\
&= \sum_{t\in\mathcal{T}} \sum_{j\in L_t} \left[D^j_{t+12} (\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt}) - \log\left(1 + e^{(\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt})}\right)\right].
\end{aligned} \tag{60}$$

B.3 Derivation of Log-likelihood of PD Model with Explicit Macroeconomic Characteristics

In this appendix, the log-likelihood for the PD model with macroeconomic characteristics as given in Equation (32) in Section 4.1.1 will be derived. Define the set of macroeconomic characteristics observed at time t as $M_t = \{M_{it} \mid i \in i, ..., k_m\}$ and define the set of macroeconomic characteristics across all periods in the set of periods $\mathcal{T}$ as $M_{\mathcal{T}} = \{M_t \mid t \in \mathcal{T}\}$. The likelihood can be written as

$$\begin{aligned}
\log L(\beta, \gamma \mid X_{\mathcal{T}}, D_{\mathcal{T}}, M_{\mathcal{T}}) &= \sum_{t \in \mathcal{T}} \sum_{j \in L_t} \Bigg[D^j_{t+12} \log \left(\frac{1}{1 + e^{-\beta_0 - \sum_{i=1}^k \beta_i X_{i,jt} - \sum_{i=1}^k \gamma_i M_{it}}} \right) \\
&\qquad + (1 - D^j_{t+12}) \log \left(\frac{e^{-\beta_0 - \sum_{i=1}^k \beta_i X_{i,jt} - \sum_{i=1}^k \gamma_i M_{it}}}{1 + e^{-\beta_0 - \sum_{i=1}^k \beta_i X_{i,jt} - \sum_{i=1}^k \gamma_i M_{it}}} \right) \Bigg] \\
&= \sum_{t \in \mathcal{T}} \sum_{j \in L_t} \Bigg[D^j_{t+12} \log \left(\frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^k \beta_i X_{i,jt} + \sum_{i=1}^k \gamma_i M_{it})}} \right) \\
&\qquad - D^j_{t+12} \log \left(\frac{e^{-(\beta_0 + \sum_{i=1}^k \beta_i X_{i,jt} + \sum_{i=1}^k \gamma_i M_{it})}}{1 + e^{-(\beta_0 + \sum_{i=1}^k \beta_i X_{i,jt} + \sum_{i=1}^k \gamma_i M_{it})}} \right) \\
&\qquad + \log \left(\frac{e^{-(\beta_0 + \sum_{i=1}^k \beta_i X_{i,jt} + \sum_{i=1}^k \gamma_i M_{it})}}{1 + e^{-(\beta_0 + \sum_{i=1}^k \beta_i X_{i,jt} + \sum_{i=1}^k \gamma_i M_{it})}} \right) \Bigg] \\
&= \sum_{t \in \mathcal{T}} \sum_{j \in L_t} \Bigg[D^j_{t+12} \log \left(e^{\beta_0 + \sum_{i=1}^k \beta_i X_{i,jt} + \sum_{i=1}^k \gamma_i M_{it}} \right) \\
&\qquad - \log \left(1 + e^{\beta_0 + \sum_{i=1}^k \beta_i X_{i,jt} + \sum_{i=1}^k \gamma_i M_{it}} \right) \Bigg] \\
&= \sum_{t \in \mathcal{T}} \sum_{j \in L_t} \left[D^j_{t+12} (\beta_0 + \sum_{i=1}^k \beta_i X_{i,jt} + \sum_{i=1}^k \gamma_i M_{it}) - \log \left(1 + e^{\beta_0 + \sum_{i=1}^k \beta_i X_{i,jt} + \sum_{i=1}^k \gamma_i M_{it}} \right) \right].
\end{aligned} \tag{61}$$

B.4 Numerical Approximation Unconditional Likelihood of the MEL PD Model

In this appendix, the Gauss-Hermite approximation of the integral in Equation (35) in Section 4.2 is provided.

$$\begin{aligned}
\log\ L(\beta, \sigma \mid X_{\mathcal{T}}, D_{\mathcal{T}}) &= \sum_{t \in \mathcal{T}} \sum_{j \in L_t} \left[\int_{-\infty}^{\infty} D^j_{t+12} (\beta_0 + \sum_{i=1}^k \beta_i X_{i,jt} + u_t^*) - \log \left(1 + e^{\beta_0 + \sum_{i=1}^k \beta_i X_{i,jt} + u_t^*} \right)\ f_{u_t}(u_t^*) du_t^* \right] \\
&= \sum_{t \in \mathcal{T}} \sum_{j \in L_t} \left[\int_{-\infty}^{\infty} D^j_{t+12} (\beta_0 + \sum_{i=1}^k \beta_i X_{i,jt} + u_t^*) - \log \left(1 + e^{\beta_0 + \sum_{i=1}^k \beta_i X_{i,jt} + u_t^*} \right)\ \frac{e^{-\frac{1}{2}\left(\frac{u_t^*}{\sigma}\right)^2}}{\sqrt{2\pi\sigma^2}} du_t^* \right] \\
&= \sum_{t \in \mathcal{T}} \sum_{j \in L_t} [\int_{-\infty}^{\infty} D^j_{t+12} (\beta_0 + \sum_{i=1}^k \beta_i X_{i,jt} + \sqrt{2\sigma^2} u_t^*) \\
&\qquad - \log \left(1 + e^{\beta_0 + \sum_{i=1}^k \beta_i X_{i,jt} + \sqrt{2\sigma^2} u_t^*} \right)\ \frac{e^{-u_t^{*2}}}{\sqrt{2\pi\sigma^2}} \sqrt{2\sigma^2} du_t^*] \\
&= \sum_{t \in \mathcal{T}} \sum_{j \in L_t} \frac{1}{\sqrt{\pi}} \left[\int_{-\infty}^{\infty} D^j_{t+12} (\beta_0 + \sum_{i=1}^k \beta_i X_{i,jt} + \sqrt{2\sigma^2} u_t^*) - \log \left(1 + e^{\beta_0 + \sum_{i=1}^k \beta_i X_{i,jt} + \sqrt{2\sigma^2} u_t^*} \right)\ e^{-u_t^{*2}} du_t^* \right] \\
&\approx \sum_{t \in \mathcal{T}} \sum_{j \in L_t} \left[\sum_{h=1}^H \frac{w_h}{\sqrt{\pi}} D^j_{t+12} (\beta_0 + \sum_{i=1}^k \beta_i X_{i,jt} + \sqrt{2\sigma^2} z_h) - \log \left(1 + e^{\beta_0 + \sum_{i=1}^k \beta_i X_{i,jt} + \sqrt{2\sigma^2} z_h} \right) \right].
\end{aligned} \tag{62}$$

B.5 Derivations PD Model with the State-of-the-Economy as a GAS Recursion

B.5.1 Log-likelihood Derivation

In this appendix, the log-likelihood objective for the PD Model with the State-of-the-Economy as a GAS Recursion in Equation (45) is derived. The derivation is similar to before, but in this appendix the likelihood will first be defined within a period t rather than across all periods t, since this expression can be used in the derivations of ∇_t and S_t. Start by writing the pdf for D^j_{t+12}. For simplicity, shorten the expressions by defining $c_{jt} = \beta_0 + \sum_{j=1}^k \beta_i X_{i,jt} + u_t$

$$D^j_{t+12} \sim p\left(D^j_{t+12} \mid \mathcal{F}_t;\theta\right) = \begin{cases} p^j_{t,t+12} = \frac{1}{1+\exp(-c_{jt})}, & \text{if } D^j_{t+12} = 1 \\ 1 - p^j_{t,t+12} = \frac{\exp(-c_{jt})}{1+\exp(-c_{jt})}, & \text{if } D^j_{t+12} = 0. \end{cases} \tag{63}$$

The likelihood L can be expressed as

$$\begin{aligned} L(\theta \mid X_t, D_{t+12}, u_t) &= \prod_{j\in L_t s.t. \{D^j_{t+12}=1\}} p\left(D^j_{t+12} \mid \mathcal{F}_t;\theta\right) * \prod_{j\in L_t s.t. \{D^j_{t+12}=0\}} \left(1 - p\left(D^j_{t+12} \mid \mathcal{F}_t;\theta\right)\right) \\ &= \prod_{j\in L_t} p\left(D^j_{t+12} \mid \mathcal{F}_t;\theta\right)^{D^j_{t+12}} \left(1 - p\left(D^j_{t+12} \mid \mathcal{F}_t;\theta\right)\right)^{(1-D^j_{t+12})} \\ \log L(\theta \mid X_t, D_{t+12}, u_t) &= \sum_{j\in L_t} D^j_{t+12} \log\left(p\left(D^j_{t+12} \mid \mathcal{F}_t;\theta\right)\right) + (1 - D^j_{t+12}) \log\left(1 - p\left(D^j_{t+12} \mid \mathcal{F}_t;\theta\right)\right). \end{aligned} \tag{64}$$

Simplifying further

$$\begin{aligned} \log L(\theta \mid X_t, D_{t+12}, u_t) &= \sum_{j\in L_t} D^j_{t+12} \log\left(\frac{1}{1+e^{-c_{jt}}}\right) + \left(1 - D^j_{t+12}\right) \log\left(\frac{e^{-c_{jt}}}{1+e^{-c_{jt}}}\right) \\ &= \sum_{j\in L_t} -D^j_{t+12} \log\left(1 + e^{-c_{jt}}\right) + \left(-D^j_{t+12} + 1\right)\left(-c_{jt} - \log\left(e^{-c_{jt}} + 1\right)\right) \\ &= \sum_{j\in L_t} -D^j_{t+12} \log\left(1 + e^{-c_{jt}}\right) - c_{jt} - \log\left(e^{-c_{jt}} + 1\right) + c_{jt} D^j_{t+12} + D^j_{t+12} \log\left(1 + e^{-c_{jt}}\right) \\ &= \sum_{j\in L_t} -c_{jt} - \log\left(e^{-c_{jt}} + 1\right) + c_{jt} D^j_{t+12} \\ &= \sum_{j\in L_t} -(\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt} + u_t) - \log\left(e^{-(\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt} + u_t)} + 1\right) + D^j_{t+12} * (\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt} + u_t). \end{aligned} \tag{65}$$

Equation (65) will be used for the derivation of ∇_t and S_t later in this appendix. To obtain the likelihood minimization objective, this equation should be summed over all periods in $\mathcal{T}$ like before. The likelihood objective is then

$$\begin{aligned} \hat{\theta} &= min_\theta - \sum_{t\in\mathcal{T}} \sum_{j\in L_t} \left[D^j_{t+12}(\beta_0 + \sum_{i=1}^k \beta_i X^j_{i,t} + u_t) - \log\left(1 + e^{\beta_0 + \sum_{i=1}^k \beta_i X^j_{i,t} + u_t}\right) \right], \text{ where} \\ u_{t+12} &= \omega + A \cdot S_t \cdot \left[\sum_{j\in L_t} D^j_{t+12} - \frac{e^{\beta_0 + \sum_{i=1}^k \beta_i X^j_{i,t} + u_t}}{e^{\beta_0 + \sum_{i=1}^k \beta_i X^j_{i,t} + u_t} + 1} \right] + B u_t, \end{aligned} \tag{66}$$

where the derivations of ∇_t and S_t are provided below.

B.5.2 Derivation of Gradient ∇_t

In this appendix, the derivation of the gradient ∇_t of the GAS model in Equation (22) is provided. The starting point is the likelihood in a given period t given by Equation (65) in the appendix above, ∇_t is then the derivative of this expression with respect to u_t. Write

$$\begin{aligned}\nabla_t &= \sum_{j\in L_t} \nabla_{j,t} = \sum_{j\in L_t} \frac{\partial \log p\left(D^j_{t+12} \mid \mathcal{F}_t;\theta\right)}{\partial u_t} = \sum_{j\in L_t} \frac{\partial \log L(\theta \mid \mathcal{F}_t)}{\partial u_t} \\ &= \sum_{j\in L_t} \underbrace{\frac{\partial\left(-(\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt} + u_t)\right)}{\partial u_t}}_{1} + \underbrace{\frac{\partial - \log\left(e^{-(\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t)} + 1\right)}{\partial u_t}}_{2} + \underbrace{\frac{\partial\left(D^j_{t+12} * (\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt} + u_t)\right)}{\partial u_t}}_{3}.\end{aligned} \tag{67}$$

Each of the three terms can be rewritten as

$$1: \quad \frac{\partial\left(-(\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt} + u_t)\right)}{\partial u_t} = -1 \tag{68}$$

$$2: \quad \frac{\partial - \log\left(e^{-(\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t)} + 1\right)}{\partial u_t} = \frac{1}{e^{\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t} + 1} \tag{69}$$

$$3: \quad \frac{\partial\left(D^j_{t+12} * (\beta_0 + \sum_{j=1}^k \beta_i X_{i,jt} + u_t)\right)}{\partial u_t} = D^j_{t+12} \tag{70}$$

Where for the second equality, the chain rule is applied as in

$$\begin{aligned}\frac{d\log(\mathrm{e}^{-x}+1)}{dx} &= \frac{1}{\mathrm{e}^{-x}+1}\cdot\frac{d(\mathrm{e}^{-x}+1)}{dx} = \frac{\frac{d\mathrm{e}^{-x}}{dx} + \frac{d1}{dx}}{\mathrm{e}^{-x}+1} = \frac{\mathrm{e}^{-x}\cdot\frac{d-x}{dx} + 0}{\mathrm{e}^{-x}+1} \\ &= \frac{\left(-\frac{dx}{dx}\right)\mathrm{e}^{-x}}{\mathrm{e}^{-x}+1} = -\frac{\mathrm{e}^{-x}}{\mathrm{e}^{-x}+1} = -\frac{1}{e^{x}+1}.\end{aligned} \tag{71}$$

To now compute ∇_t, the sum of Equation (68), Equation (69) and Equation (70) can now be substituted in Equation (67). Thus, ∇_t can be written as

$$\begin{aligned}\nabla_t &= \sum_{j\in L_t} \nabla_{j,t} = \sum_{j\in L_t} \frac{\partial \log p\left(D^j_{t+12} \mid u_t, X^t, D^j_{t-1};\theta\right)}{\partial u_t} \\ &= \sum_{j\in L_t} -1 + \frac{1}{e^{\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t} + 1} + D^j_{t+12} \\ &= \sum_{j\in L_t} D^j_{t+12} - \frac{e^{\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t}}{e^{\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t} + 1}.\end{aligned} \tag{72}$$

B.5.3 Derivation of Scaling Function S_t

In Equation (23) in Section 3.2.4, the scaling function S_t that is used within the GAS model is defined. In this appendix, the derivation for this scaling is provided. Recall that the scaling matrix is defined by

$$S_t = \mathcal{I}_{t-1}^{-d} = -E_{t-1}\left[\frac{\partial^2 \log p\left(D^j_{t+12} \mid \mathcal{F}_t;\theta\right)}{\partial^2 u_t}\right]^{-d}. \tag{73}$$

Given that

$$\nabla_t = \sum_{j\in L_t} \frac{\partial \log p\left(D^j_{t+12} \mid \mathcal{F}_t;\theta\right)}{\partial u_t} = \sum_{j\in L_t} D^j_{t+12} - \frac{e^{\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t}}{e^{\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t} + 1}. \tag{74}$$

It also follows that

$$\frac{\partial \log p\left(D^j_{t+12} \mid \mathcal{F}_t;\theta\right)}{\partial u_t} = D^j_{t+12} - \frac{e^{\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t}}{e^{\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t} + 1}. \tag{75}$$

Continue by writing

$$\begin{aligned}
\frac{\partial^2 \log p\left(D^j_{t+12} \mid \mathcal{F}_t;\theta\right)}{\partial^2 u_t} &= \frac{\partial}{\partial u_t}\frac{\partial \log p\left(D^j_{t+12} \mid \mathcal{F}_t;\theta\right)}{\partial u_t} \\
&= \frac{\partial D^j_{t+12}}{\partial u_t} - \frac{\partial}{\partial u_t}\left(\frac{e^{\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t}}{e^{\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t}+1}\right) \\
&= 0 - \frac{e^{\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t}\left(e^{\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t}+1\right) - e^{\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t}e^{\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t}}{\left(e^{\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t}+1\right)^2} \\
&= -\frac{e^{\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t}}{\left(e^{\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t}+1\right)^2}.
\end{aligned} \tag{76}$$

So S_t in Equation (73) can be written as

$$S_t = -E_{t-1}\left[\frac{\partial^2 \log p\left(D^j_{t+12} \mid \mathcal{F}_t;\theta\right)}{\partial^2 u_t}\right]^{-d} = -E_{t-1}\left[-\frac{e^{\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t}}{\left(e^{\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t}+1\right)^2}\right]^{-d} = E_{t-1}\left[\frac{e^{\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t}}{\left(e^{\beta_0+\sum_{j=1}^k \beta_i X_{i,jt}+u_t}+1\right)^2}\right]^{-d}. \tag{77}$$

C Supplementary Tables and Figures

C.1 Regression Results of Logistic Regression Including Recovered Loan Dummy

In Section 3.1, the assumption that loans having defaulted are statistically different from those that have not was proposed. In this appendix, some further substance to this assumption is provided via Table 7. In this table, the coefficient on Recovered Loan is significant at any conventional conifdence threshold, and it is therefore reasonable to assume that loans that have defaulted before are statistically different from those that have not.

Risk Driver	**Coefficient**	**Std.Err.**	**t-statistic**
Recovered Loan	0.8186	0.0453	18.0706
Intercept	-3.3966	0.0954	-35.5932
Loan Age	0.0030	0.0001	24.4258
Current Deferred UPB	-0.00002	0.0001	0.2120
Credit Score	-0.0100	0.0001	-113.7807
Mortgage Insurance Percentage	0.0021	0.0006	3.3807
Number Of Units	-0.0327	0.0209	-1.5671
Original Debt To Income Ratio	0.0240	0.0005	49.0548
Original UPB	3.1767	0.0540	58.813
Original Loan To Value	0.0224	0.0005	40.8968
Original Interest Rate	0.2871	0.0056	51.1811
Loan Term	0.0038	0.0001	36.6835
Number Of Borrowers	0.6260	0.0108	57.7615
First Time Homebuyer Flag Y	0.1260	0.0175	7.2125
Channel C	-0.1351	0.0280	-4.8200
Channel R	-0.1563	0.0253	-6.1759
Channel T	0.0023	0.0262	0.0863
Loan Purpose N	-0.3936	0.0140	-28.1091
Loan Purpose P	-0.5744	0.0149	-38.6205

Table 7: *Estimation Results Benchmark Model with Recovered Loan Dummy*

C.2 Supplementary Tables of the Freddie Mac Data

In this appendix, additional summary statistics related to the Freddie Mac mortgage data, as introduced in Section 5, are provided. These statistics are intended to enhance the understanding of the Freddie Mac data set.

Risk Driver	Mean	Std	min	50%	max
Credit Score	739.15	52.89	300	749	850
Current Deferred UPB	107.39	2592.26	0	0	306100
Loan Age	38.42	38.56	0	25	277
Mortgage Insurance Percentage	4.43	9.85	0	0	55
Original Debt-to-Income Ratio	33.42	11.28	1	34	65
Original Interest Rate	5.15	1.32	1.75	5.12	12
Original Loan-to-Value Ratio	70.30	17.53	3	75	105
Original Unpaid Balance	182772	110640	8000	155000	1867000
Loan Term	306.279	81.39	48	360	480
Total No. Observations	3817151				

Table 8: *Summary Statistics for Continuous Variables in the Freddie Mac Data Set*

Risk Driver				
Loan Purpose	P (39%)	N (32%)	C (29%)	
First Time Homebuyer	Yes (12%)	No (88%)		
No. Units	1 (97%)	2 (1.7%)	3 (0.3%)	4 (0.2%)
No. Borrowers	1 (55%)	>1 (45%)		
Channel	R (53%)	N.S. (23%)	C (17%)	B (5.6%)

Table 9: *Summary Statistics for Categorical Variables in the Freddie Mac Data Set*

C.2.1 Summary Statistics: Default Indicator

In Figure 9 the evolution of the default rate over time is shown. In Table 10 thereunder, the same data is shown, but also including the absolute number of details in each year to illustrate the scale the estimation at which the estimations are performed.

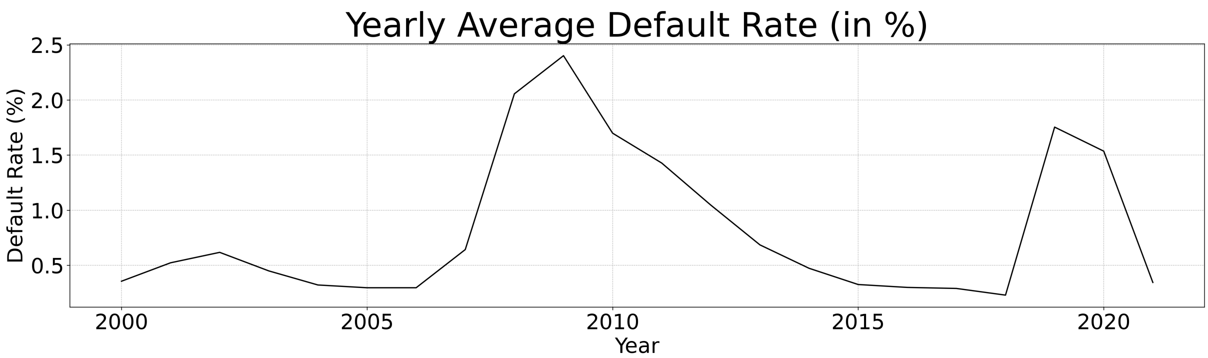


Figure 9: *The observed annual default rate plotted over time in %. Note that in this figure, the outcome variable D^j_{t+12} is directly plotted, so rather than the actual moments of default, each the indicator indicates whether the observation will be in default in 12 months. This is done because this is also the target quantity in the estimation process, but therefore it might seem that the spike around the COVID-19 crisis and the spike around the financial crisis are too "early".*

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Default Rate	0.35%	0.52%	0.62%	0.45%	0.32%	0.30%	0.30%	0.64%	2.05%	2.40%	1.70%
No. Defaults	70	129	143	117	109	127	152	383	1307	1570	1045

Year	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Default Rate	1.43%	1.05%	0.68%	0.47%	0.33%	0.29%	0.29%	0.23%	1.75%	1.54%	0.34%
No. Defaults	811	517	341	258	192	189	201	174	1272	947	212

Table 10: *Annual Defaults: Relative and Absolute*

C.3 Supplementary Figures of the Macroeconomic Data

C.3.1 Correlations of Macroeconomic Quantities with Mortgage Defaults

In this appendix, the auto-correlation of the macroeconomic quantities up to a lag of 24 months is reported. In Figure 10 below, the lag with the largest correlation is marked by the red dot.

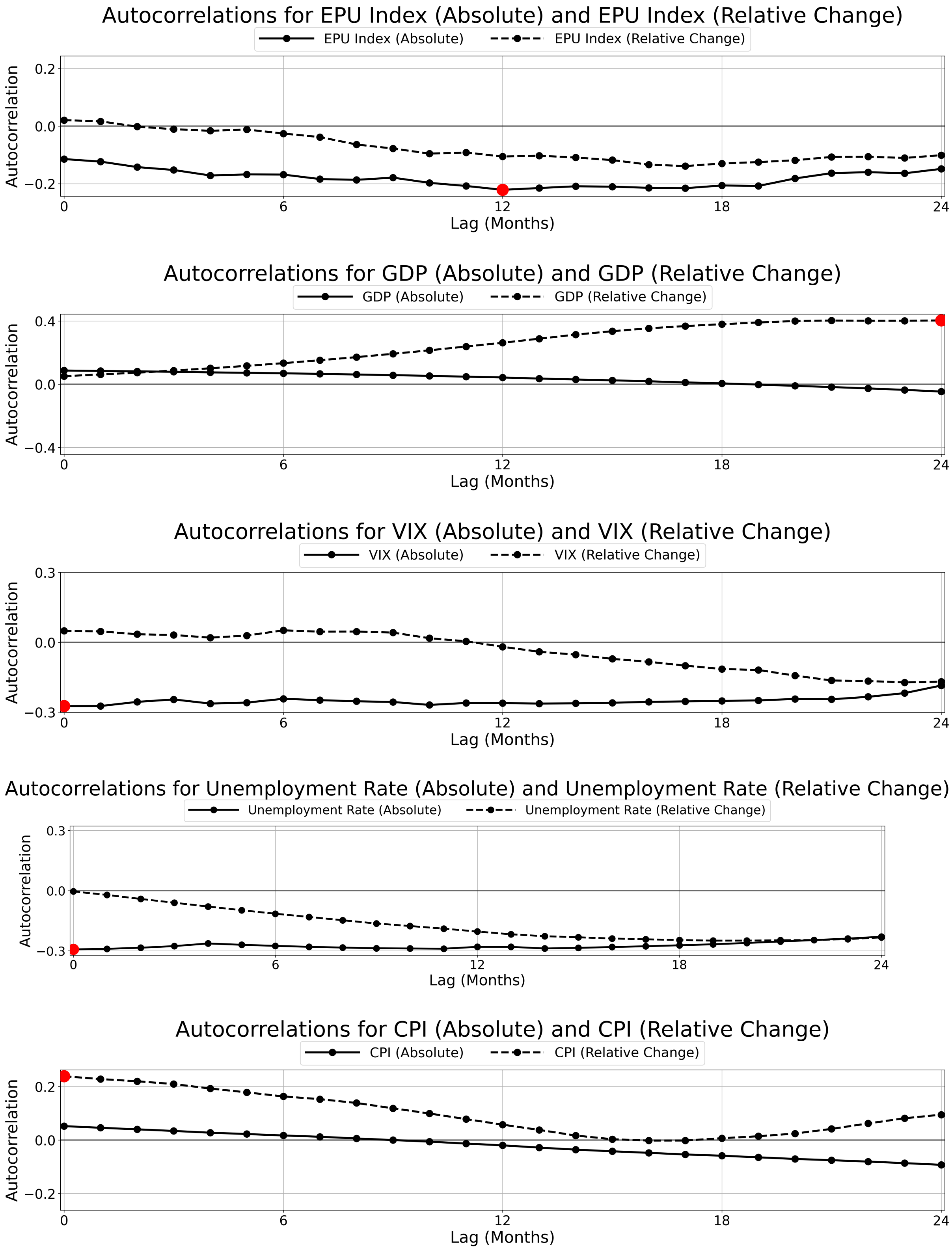


Figure 10: *Visualization of Autocorrelation of Macroeconomic Variables for Various Lags*

C.3.2 Overlay of Macroeconomic Quantities and Observed Default Rate

In this appendix, the ODR is overlaid with each of the raw macroeconomic risk drivers for easy verification of the associated correlation. What is immediately visible is that for some of the risk drivers, with appropriate scaling, bear some resemblance with the ODR.

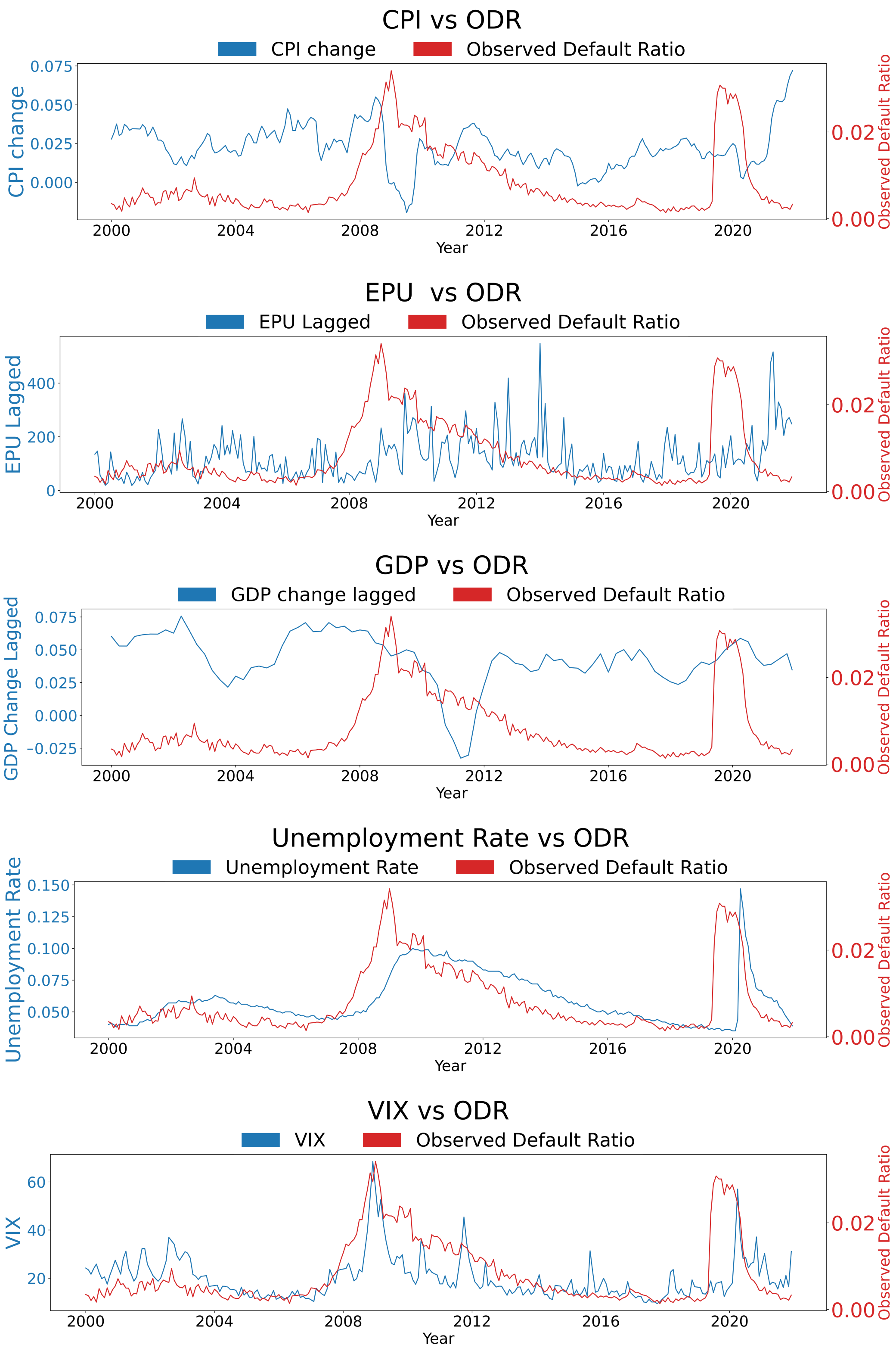


Figure 11: *Visualization of Correlation of Macroeconomic Variables with Average Loan Default*

C.4 Supplementary Tables for Section 6

In this appendix, the results that are reported in Section 6.1 are discussed in further detail. In particular, instead of reporting exclusively the weighted average of the performance measures, this appendix aims to expand on these metrics by also reporting them from year-to-year. The weights used to compute the weighted average are determined by the number of observations in that given year in the relevant set divided by the total number of observations. These are provided in Table 11 below.

Year	No. Obs. Testing Set	No. Obs.Training Set	No. Obs. Total
2000	19808	46219	66027
2001	24671	57566	82237
2002	23105	53912	77017
2003	25958	60569	86527
2004	33978	79282	113260
2005	42795	99855	142650
2006	51305	119712	171017
2007	59593	139051	198644
2008	63590	148377	211967
2009	65346	152474	217820
2010	61501	143503	205004
2011	56761	132443	189204
2012	49451	115385	164837
2013	49806	116213	166020
2014	54485	127131	181617
2015	59034	137745	196780
2016	63133	147310	210444
2017	69207	161482	230690
2018	75729	176703	252430
2019	72499	169164	241664
2020	61674	143905	205580
2021	61717	144007	205724
Cumulative	1145146	2672014	3817151

Table 11: *Total Number of Observations in Each Year. Note: In the OOT procedure, only the No. Obs. Total is relevant as within years, data are not split in train test sets.*

C.4.1 Supplementary Tables OOS - Training Set

In this Appendix, the OOS results from the training set in Table 4 in Section 6.1 are reported in more detail. In the table, the overall AUC or MSE is reported, but not the individual contributions. This appendix will provide further insights into the year-to-year performance of the training set of the OOS testing procedure.

In order, in this appendix the AUC results, individual MSE results and the results of the MSE of the yearly averages are reported in Table 12, Table 13 and Table 14.

Year	AUC IC	AUC Benchmark	AUC Macro	AUC GAS	AUC MEL
2000	0.5	0.8615	0.8610	0.8657	0.8664
2001	0.5	0.8476	0.8462	0.8671	0.8568
2002	0.5	0.8496	0.8471	0.8513	0.8593
2003	0.5	0.8720	0.8712	0.8724	0.8910
2004	0.5	0.8568	0.8764	0.8585	0.8714
2005	0.5	0.8347	0.8356	0.8385	0.8468
2006	0.5	0.8106	0.8527	0.8178	0.8164
2007	0.5	0.8094	0.8122	0.8059	0.8437
2008	0.5	0.8124	0.8081	0.8101	0.8174
2009	0.5	0.8129	0.8118	0.8079	0.8193
2010	0.5	0.8007	0.7981	0.7916	0.8116
2011	0.5	0.8026	0.7970	0.7937	0.8126
2012	0.5	0.8175	0.8155	0.8138	0.8258
2013	0.5	0.8254	0.8235	0.8096	0.8357
2014	0.5	0.8310	0.8272	0.8282	0.8398
2015	0.5	0.8053	0.8026	0.8045	0.8110
2016	0.5	0.7855	0.7840	0.7996	0.7738
2017	0.5	0.7684	0.7894	0.7882	0.7839
2018	0.5	0.7648	0.7612	0.7765	0.7627
2019	0.5	0.7393	0.7529	0.7560	0.7577
2020	0.5	0.7750	0.7496	0.7487	0.7575
2021	0.5	0.7592	0.7659	0.7593	0.7790
W. Avg.	0.5	0.8023	0.8136	0.8091	0.8376

Table 12: *Yearly AUC Results - OOS Train*

Year	MSE IC	MSE Benchmark	MSE Macro	MSE GAS	MSE MEL
2000	0.0036	0.0037	0.0036	0.0036	0.0042
2001	0.0051	0.0051	0.0050	0.0048	0.0050
2002	0.0059	0.0059	0.0059	0.0064	0.0058
2003	0.0047	0.0047	0.0047	0.0045	0.0046
2004	0.0031	0.0031	0.0031	0.0032	0.0030
2005	0.0030	0.0030	0.0030	0.0031	0.0029
2006	0.0031	0.0032	0.0031	0.0032	0.0030
2007	0.0066	0.0066	0.0065	0.0064	0.0065
2008	0.0201	0.0195	0.0194	0.0197	0.0193
2009	0.0234	0.0227	0.0224	0.0226	0.0223
2010	0.0165	0.0162	0.0162	0.0164	0.0162
2011	0.0141	0.0139	0.0139	0.0141	0.0139
2012	0.0103	0.0102	0.0102	0.0103	0.0102
2013	0.0069	0.0068	0.0068	0.0069	0.0068
2014	0.0048	0.0048	0.0048	0.0046	0.0047
2015	0.0033	0.0033	0.0033	0.0033	0.0033
2016	0.0029	0.0030	0.0029	0.0030	0.0029
2017	0.0028	0.0028	0.0028	0.0029	0.0028
2018	0.0022	0.0022	0.0022	0.0023	0.0022
2019	0.0173	0.0172	0.0173	0.0172	0.0171
2020	0.0153	0.0152	0.0152	0.0152	0.0151
2021	0.0034	0.0035	0.0034	0.0035	0.0034
W. Avg.	0.0088	0.0087	0.0087	0.0088	0.0087

Table 13: *Yearly MSE Results - OOS Train*

Year	A. MSE IC	A. MSE Benchm.	A. MSE Macro	A. MSE GAS	A. MSE MEL
2000	2.89E-05	0.00011	4.46E-05	9.08E-05	3.54E-04
2001	1.47E-05	6.46E-05	3.53E-05	2.79E-05	4.79E-06
2002	8.96E-06	3.82E-05	3.9E-05	2.91E-05	2.07E-06
2003	1.77E-05	1.76E-05	1.49E-05	1.42E-05	6.21E-07
2004	3.44E-05	2.72E-05	1.11E-05	1.85E-05	1.51E-07
2005	3.6E-05	3.42E-05	8.69E-06	2.31E-05	3.56E-08
2006	3.46E-05	5.02E-05	1.1E-05	2.49E-05	1.26E-07
2007	5.25E-06	2.36E-05	2.24E-06	1.66E-06	9.69E-06
2008	0.000131	6.53E-05	3.34E-05	0.000117	1.92E-04
2009	0.000219	0.000152	6.55E-06	2.28E-05	5.8E-05
2010	6.08E-05	3.49E-05	4.28E-06	4.52E-07	1.38E-05
2011	2.89E-05	1.66E-05	9.19E-07	1.09E-06	5.69E-07
2012	2.22E-06	1.3E-06	5.03E-07	4.62E-07	2.07E-06
2013	4.11E-06	1.58E-06	2.81E-07	5.89E-06	2.43E-06
2014	1.73E-05	7.94E-06	3.25E-06	8E-06	1.57E-06
2015	3.21E-05	1.49E-05	7.07E-06	1.11E-05	8.71E-07
2016	3.64E-05	1.58E-05	4.83E-06	6.51E-06	2.4E-08
2017	3.79E-05	1.66E-05	2.67E-06	5.47E-06	2.76E-10
2018	4.6E-05	2.44E-05	8.34E-06	7.72E-06	4.41E-07
2019	7.37E-05	0.000106	0.000155	0.000159	2.37E-04
2020	4.27E-05	7.54E-05	2E-05	3.51E-05	1.6E-08
2021	3.04E-05	8.12E-06	6.15E-06	1.57E-05	9.67E-05
W. Avg.	4.88E-05	4.09E-05	1.92E-05	2.88E-05	4.21E-05

Table 14: *Annual MSE results for OOS - Train*

C.4.2 Supplementary Tables OOS - Testing Set

In this Appendix, the OOS results from the training set in Table 4 and in Figure 6 in Section 6.1 are reported in more detail. In the table and figures, the overall AUC or MSE are reported, but not the individual contributions. This appendix will provide further insight into the year-to-year performance of the testing set of the OOS testing procedure.

In order, in this appendix the AUC results, individual MSE results and the results of the MSE of the yearly averages are reported in Table 15, Table 16 and Table 17.

Year	AUC IC	AUC Benchmark	AUC Macro	AUC GAS	AUC MEL
2000	0.5	0.8198	0.8178	0.8342	0.8299
2001	0.5	0.8519	0.8535	0.8321	0.8631
2002	0.5	0.8519	0.8508	0.8512	0.8057
2003	0.5	0.8844	0.8815	0.8711	0.8938
2004	0.5	0.8670	0.8660	0.8636	0.8780
2005	0.5	0.8171	0.8183	0.8324	0.8369
2006	0.5	0.7976	0.7991	0.8179	0.8009
2007	0.5	0.8165	0.8193	0.8266	0.8219
2008	0.5	0.8127	0.8069	0.8081	0.8175
2009	0.5	0.8105	0.8068	0.8055	0.8197
2010	0.5	0.8017	0.7983	0.7990	0.8094
2011	0.5	0.8014	0.7952	0.7830	0.8100
2012	0.5	0.8165	0.8155	0.7961	0.8260
2013	0.5	0.7949	0.7920	0.8098	0.8091
2014	0.5	0.8383	0.8397	0.8192	0.8461
2015	0.5	0.8079	0.8021	0.8240	0.8203
2016	0.5	0.8102	0.8065	0.8310	0.8136
2017	0.5	0.7806	0.8208	0.8414	0.7862
2018	0.5	0.7825	0.8088	0.7871	0.7880
2019	0.5	0.7391	0.8275	0.7791	0.7235
2020	0.5	0.7470	0.7825	0.7651	0.7595
2021	0.5	0.7455	0.7725	0.7751	0.7533
W. Avg.	0.5	0.8006	0.8137	0.8086	0.8044

Table 15: *AUC results Individual PDs - OOS Test*

Year	MSE IC	MSE Benchmark	MSE Macro	MSE GAS	MSE MEL
2000	0.0034	0.0036	0.0034	0.0036	0.0042
2001	0.0051	0.0051	0.0051	0.0059	0.0050
2002	0.0062	0.0062	0.0062	0.0053	0.0061
2003	0.0037	0.0037	0.0037	0.0043	0.0036
2004	0.0034	0.0034	0.0034	0.0034	0.0033
2005	0.0029	0.0029	0.0029	0.0030	0.0028
2006	0.0028	0.0029	0.0028	0.0031	0.0027
2007	0.0061	0.0061	0.0061	0.0067	0.0061
2008	0.0201	0.0195	0.0194	0.0198	0.0193
2009	0.0237	0.0229	0.0226	0.0227	0.0225
2010	0.0172	0.0168	0.0168	0.0168	0.0168
2011	0.0140	0.0138	0.0139	0.0137	0.0138
2012	0.0104	0.0103	0.0103	0.0103	0.0104
2013	0.0066	0.0066	0.0066	0.0068	0.0066
2014	0.0045	0.0045	0.0045	0.0050	0.0045
2015	0.0032	0.0032	0.0032	0.0033	0.0032
2016	0.0032	0.0032	0.0032	0.0030	0.0032
2017	0.0031	0.0031	0.0030	0.0029	0.0030
2018	0.0027	0.0027	0.0027	0.0023	0.0026
2019	0.0170	0.0169	0.0169	0.0171	0.0169
2020	0.0149	0.0148	0.0136	0.0150	0.0147
2021	0.0033	0.0034	0.0033	0.0034	0.0033
W. Avg.	0.0088	0.0087	0.0087	0.0088	0.0087

Table 16: *MSE results Individual PDs - OOS Test*

Year	A. MSE IC	A. MSE Benchm.	A. MSE Macro	A. MSE GAS	A. MSE MEL
2000	3.10E-05	1.14E-04	4.72E-05	8.97E-05	3.70E-04
2001	1.45E-05	6.32E-05	3.50E-05	2.97E-05	3.37E-06
2002	7.24E-06	3.47E-05	3.54E-05	2.91E-05	5.50E-06
2003	2.74E-05	2.81E-05	2.38E-05	1.30E-05	3.35E-07
2004	3.10E-05	2.36E-05	9.16E-06	1.85E-05	3.41E-06
2005	3.74E-05	3.51E-05	9.40E-06	2.19E-05	1.89E-06
2006	3.85E-05	5.49E-05	1.32E-05	2.46E-05	1.24E-07
2007	7.68E-06	2.76E-05	3.91E-06	1.65E-06	1.21E-05
2008	1.32E-04	6.48E-05	3.34E-05	1.19E-04	1.93E-04
2009	2.27E-04	1.54E-04	7.94E-06	2.35E-05	4.61E-05
2010	7.13E-05	4.42E-05	7.37E-06	5.88E-07	1.29E-05
2011	2.77E-05	1.58E-05	7.17E-07	1.18E-06	1.99E-07
2012	2.59E-06	1.67E-06	6.87E-07	2.93E-07	2.56E-06
2013	5.27E-06	1.90E-06	6.34E-07	6.30E-06	3.52E-06
2014	1.93E-05	9.53E-06	4.18E-06	7.59E-06	2.70E-06
2015	3.29E-05	1.50E-05	7.43E-06	1.09E-05	1.06E-06
2016	3.29E-05	1.26E-05	3.60E-06	6.76E-06	2.84E-07
2017	3.48E-05	1.42E-05	1.89E-06	5.53E-06	6.80E-07
2018	3.96E-05	1.98E-05	5.77E-06	7.88E-06	6.29E-06
2019	6.83E-05	9.98E-05	1.48E-04	1.60E-04	2.08E-04
2020	3.76E-05	6.93E-05	1.65E-05	3.53E-05	6.54E-08
2021	3.15E-05	8.73E-06	6.68E-06	1.59E-05	9.09E-05
W. Avg.	4.91E-05	4.09E-05	1.89E-05	2.90E-05	4.05E-05

Table 17: *Annual MSE result - OOS Test*

C.4.3 Supplementary Tables OOT - One-Year

In this Appendix, the OOT results set in Figure 7 in Section 6.2.1 are reported in more detail. In Figure 7, the overall AUC or MSE are reported, but not the individual contributions per year. This appendix will provide further insight into the year-to-year performance of the testing sets of the 1-year OOT procedure.

In order, in this appendix the AUC results, individual MSE results and the results of the MSE of the yearly averages are reported in Table 18, Table 19 and Table 20.

Year	AUC Intercept	AUC Macro	AUC Macro	AUC GAS	AUC MEL
2011	0.5	0.8029	0.7988	0.7905	0.8157
2012	0.5	0.8234	0.8175	0.8085	0.8274
2013	0.5	0.8271	0.8222	0.8096	0.8306
2014	0.5	0.8422	0.8403	0.8253	0.8443
2015	0.5	0.8171	0.8150	0.8014	0.8172
2016	0.5	0.8025	0.7989	0.7940	0.8032
2017	0.5	0.7849	0.7741	0.7800	0.7837
2018	0.5	0.7742	0.7709	0.7797	0.7716
2019	0.5	0.7201	0.7150	0.7369	0.7141
2020	0.5	0.6929	0.7424	0.7476	0.7558
2021	0.5	0.7548	0.7510	0.7640	0.7700
W. Avg.	0.5	0.7821	0.7824	0.7829	0.7900

Table 18: *AUC results Individual PDs - One-Year OOT*

Year	MSE IC	MSE Benchmark	MSE Macro	MSE GAS	MSE MEL
2011	0.0141	0.0140	0.0140	0.0140	0.0140
2012	0.0104	0.0104	0.0103	0.0103	0.0104
2013	0.0068	0.0069	0.0067	0.0068	0.0068
2014	0.0047	0.0048	0.0047	0.0047	0.0047
2015	0.0033	0.0034	0.0032	0.0033	0.0033
2016	0.0030	0.0031	0.0030	0.0030	0.0030
2017	0.0029	0.0029	0.0028	0.0029	0.0029
2018	0.0023	0.0024	0.0023	0.0023	0.0023
2019	0.0202	0.0172	0.0173	0.0172	0.0173
2020	0.0152	0.0151	0.0152	0.0154	0.0150
2021	0.0034	0.0034	0.0035	0.0034	0.0037
W. Avg.	0.0079	0.0076	0.0075	0.0075	0.0075

Table 19: *MSE results Individual PDs - One-Year OOT*

Year	A. MSE IC	A. MSE Benchm.	A. MSE Macro	A. MSE GAS	A. MSE MEL
2011	1.17E-05	5.76E-07	2.70E-06	7.72E-06	3.06E-07
2012	5.79E-07	4.52E-06	3.06E-06	1.78E-06	1.84E-06
2013	1.87E-05	1.44E-05	5.18E-06	3.61E-06	2.03E-06
2014	3.72E-05	1.90E-05	2.59E-06	6.91E-06	1.13E-06
2015	5.01E-05	1.86E-05	1.47E-06	7.27E-06	8.31E-07
2016	4.59E-05	1.17E-05	8.58E-07	4.21E-06	3.59E-08
2017	4.06E-05	7.92E-06	1.33E-06	3.51E-06	5.05E-10
2018	4.13E-05	8.88E-06	2.59E-07	4.91E-06	2.68E-07
2019	8.48E-05	1.61E-04	2.40E-04	1.80E-04	2.32E-04
2020	4.19E-05	8.18E-05	5.53E-06	2.87E-04	2.50E-09
2021	3.42E-05	1.11E-05	1.87E-06	1.14E-05	9.09E-05
W. Avg.	3.91E-05	3.31E-05	3.01E-05	5.06E-05	3.35E-05

Table 20: *Annual MSE results - One-Year OOT*

C.4.4 Supplementary Tables OOT - Three-Year

In this Appendix, the OOT results set in Figure 8 in Section 8 are reported in more detail. In Figure 8, the overall AUC or MSE are reported, but not the individual contributions per year. This appendix will provide further insight into the year-to-year performance of the testing sets of the one-year OOT procedure.

In order, in this appendix the AUC results, individual MSE results and the results of the MSE of the yearly averages are reported in Table 21, Table 22 and Table 23.

Year	AUC IC	AUC Benchmark	AUC Macro	AUC GAS	AUC MEL
2013	0.5	0.8179	0.8242	0.8096	0.8291
2014	0.5	0.8375	0.8363	0.8253	0.8420
2015	0.5	0.8122	0.8111	0.8014	0.8136
2016	0.5	0.7963	0.7988	0.7940	0.7978
2017	0.5	0.7688	0.7829	0.7800	0.7806
2018	0.5	0.7670	0.7704	0.7797	0.7701
2019	0.5	0.7132	0.7178	0.7369	0.7109
2020	0.5	0.7352	0.6386	0.7476	0.7471
2021	0.5	0.7479	0.7591	0.7640	0.7638
W. Avg.	0.5	0.7740	0.7681	0.7799	0.7803

Table 21: *AUC Results Individual PDs - Three-Year OOT*

Year	MSE IC	MSE Benchmark	MSE Macro	MSE GAS	MSE MEL
2013	0.0068	0.0070	0.0067	0.0068	0.0069
2014	0.0047	0.0050	0.0047	0.0047	0.0048
2015	0.0033	0.0035	0.0032	0.0033	0.0033
2016	0.0030	0.0032	0.0030	0.0030	0.0030
2017	0.0029	0.0030	0.0028	0.0029	0.0029
2018	0.0023	0.0024	0.0023	0.0023	0.0023
2019	0.0172	0.0172	0.0173	0.0202	0.0173
2020	0.0152	0.0152	0.0160	0.0154	0.0152
2021	0.0034	0.0034	0.0034	0.0034	0.0034
W. Avg.	0.0076	0.0067	0.0067	0.0070	0.0072

Table 22: *MSE Results Individual PDs - Three-Year OOT*

Year	A. MSE IC	A. MSE Benchm.	A. MSE Macro	A. MSE GAS	A. MSE MEL
2013	1.62E-05	2.68E-05	8.14E-06	3.61E-06	7.11E-06
2014	4.24E-05	3.64E-05	4.72E-06	6.91E-06	9.69E-06
2015	6.26E-05	3.68E-05	2.98E-06	7.27E-06	7.41E-06
2016	6.12E-05	2.55E-05	1.90E-06	4.21E-06	1.63E-06
2017	5.58E-05	1.71E-05	2.14E-06	3.51E-06	2.66E-07
2018	5.57E-05	1.62E-05	6.52E-07	4.91E-06	8.72E-08
2019	6.72E-05	0.00013	0.00025	0.00018	0.00021
2020	4.39E-05	0.00011	3.53E-08	0.00005	0.00017
2021	2.33E-05	5.05E-07	4.54E-08	1.14E-05	2.74E-06
W. Avg.	4.89E-05	4.56E-05	3.45E-05	3.31E-05	4.85E-05

Table 23: *Annual MSE results - Three-Year OOT*

C.5 Plots State-of-the-Economy Estimates vs ODR

In this appendix, the estimated state-of-the-economy, denoted as u_t, is presented for three models: the MEL model, the GAS model, and the model incorporating macroeconomic variables. These estimates are displayed in the figures below. On the left axis of each figure, the u_t values are plotted. To facilitate comparison among the three methods used for estimating u_t, a standardization process is applied. Each vector of u_t's is demeaned and is then divided by its standard deviation. The observed default rate is plotted on the right axis of the figures.

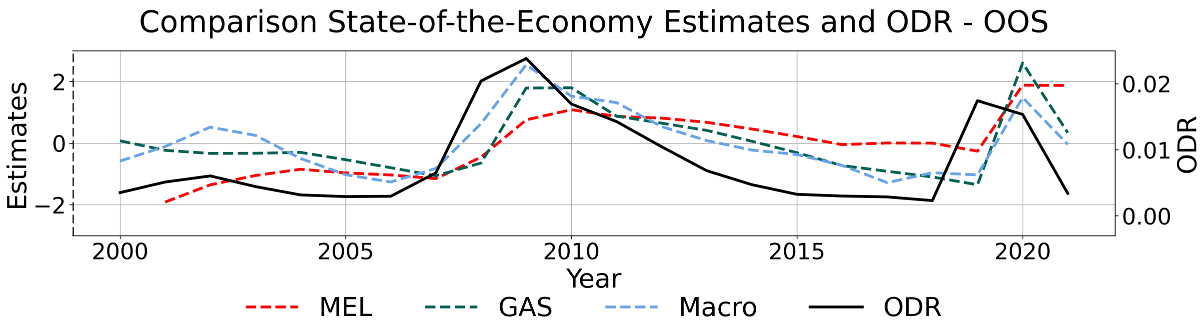


Figure 12: *Comparison of the Estimated State-of-the-Economy Versus the Observed Default Rate Based on the Out-of-Sample Estimates.*

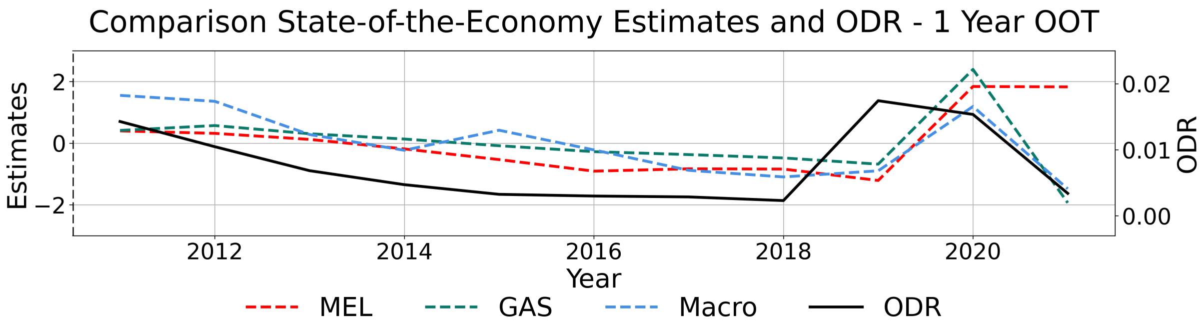


Figure 13: *Comparison of the Estimated State-of-the-Economy Versus the Observed Default Rate Based on the One-year Out-Of-Time Estimates.*

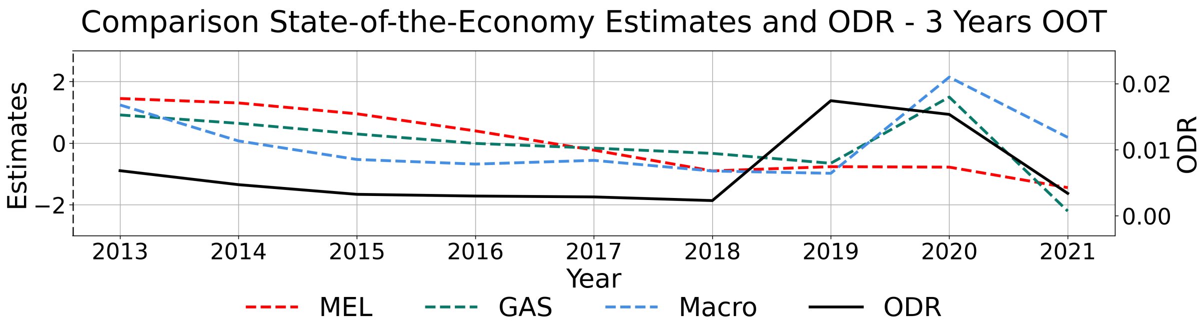


Figure 14: *CComparison of the Estimated State-of-the-Economy Versus the Observed Default Rate Based on the Three-Year Out-Of-Time Estimates.*

D Variable Selection Methods

In Section 4.4, it was discussed that the variable selection procedure in this thesis shall rely on forwards and backwards stepwise regression based on the AIC. In this appendix, the algorithm for both methods is provided.

D.1 Variable Selection by Forward Selection

Forward selection is a stepwise regression procedure starting with no variables in the model and iteratively adding the most significant variable that improves model performance, assessed using criteria the Akaike Information Criterion in this case. This iterative procedure continues until adding new variables does not lead to better model performance.

Algorithm 1: Stepwise Regression Using Forward Selection with AIC

Input: Risk Drivers X, Default Indicator D
Output: Selected features

Initialize an empty set of selected features: $S = \varnothing$
while *True* **do**
  Calculate the AIC of the current model using Equation (50)
  Initialize variables to keep track of the feature to add and the maximum AIC decrease: $x_{\text{add}} = \varnothing$, $\Delta\text{AIC}_{\text{max}} = 0$
  for *each feature x_i not in S* **do**
    Create a temporary model by adding feature x_i to the current model
    Calculate the AIC of the temporary model
    Calculate the decrease in AIC: ΔAIC = AIC of current model – AIC of temporary model
    if $\Delta AIC > \Delta AIC_{max}$ **then**
      Set $x_{\text{add}} \leftarrow x_i$ and $\Delta\text{AIC}_{\text{max}} = \Delta\text{AIC}$
    end
  end
  if $\Delta AIC_{max} > 0$ **then**
    Add the feature x_{add} to S
  end
  else
    Break // No further addition improves AIC
  end
end
return *Selected features: S*

D.2 Variable Selection by Backward Selection

Backward Selection begins with a model that includes all candidate variables and iteratively removes the least significant variable that enhances model performance. The process is again guided by the AIC.. This iterative process continues until the removal of additional variables no longer yields a performance improvement in the model.

Algorithm 2: Stepwise Regression Using Backward Selection with AIC

Input: Risk Drivers X, Default Indicator D
Output: Selected features

Initialize a set of all features: $S = \{x_1, x_2, \ldots, x_K\}$
while *True* **do**
  Calculate the AIC of the current model using Equation (50)
  Initialize variables to keep track of the feature to remove and the maximum AIC increase: $x_{\text{remove}} = \varnothing$, $\Delta\text{AIC}_{\text{max}} = 0$
  for *each feature* x_i *in* S **do**
    Create a temporary model by removing feature x_i from the current model
    Calculate the AIC of the temporary model
    Calculate the increase in AIC: ΔAIC = AIC of temporary model − AIC of current model
    if $\Delta AIC > \Delta AIC_{max}$ **then**
      Set $x_{\text{remove}} \leftarrow x_i$ and $\Delta\text{AIC}_{\text{max}} = \Delta\text{AIC}$
    end
  end
  if $\Delta AIC_{max} > 0$ **then**
    Remove the feature x_{remove} from S
  end
  else
    Break // No further removal improves AIC
  end
end
return *Selected features:* S