Utrecht University

MSc Mathematical Sciences

Master Thesis

Top Down Stress Testing:

An Application of Adaptive Lasso to Forecasting Credit Loss Rates

Author:
Tineke Blom

Supervisors:
Dr. André Ploegh (Rabobank)
Ir. Martin van Jole (Rabobank)
Dr. Cristian Spitoni (UU)
Prof. Dr. Roberto Fernández (UU)

Tuesday 17th November, 2015

Utrecht University
Faculty of Science
Master Mathematical Sciences
Princetonplein 5
3584 CC Utrecht

Rabobank Nederland
Risk Management
Integrated Risk Capital Planning and Modelling
P.O. Box 17100
3500 HG Utrecht

Abstract

The aim of this thesis is to determine the data requirements and feasibility of data-driven top-down stress testing for credit loss rates. To that end, we use the Adaptive Lasso method to simultaneously select and estimate parsimonious linear models from a very large set of potential model specifications. Adaptive Lasso is a penalized regression method which can accurately and uniquely select substantially relevant predictors and has attractive asymptotic properties. The selected models are able to give accurate forecasts in baseline and severely adverse macro-economic scenarios for the United States. We find that the loan data needs to be divided into a minimum of five categories to adequately capture the link between the macro-economy and credit loss rates. For reliable forecasts, roughly 20 years of credit loss data is required, or at least one complete business cycle must be present in the data.

Preface

This thesis was submitted in partial fulfillment of the requirements for the master Mathematical Sciences at Utrecht University. During my master's studies I developed curiosity as to how mathematical theory could be applied in practice. At Rabobank, I had the opportunity of conducting my research as an intern for the Capital Planning and Modelling team. I am thankful for the unforgettable experience and the knowledge that I gained from the experts there. I take this opportunity to express gratitude to all of the staff at Rabobank who directly or indirectly helped me to complete this thesis.

I would like to express my sincere gratitude to my supervisors Dr. André Ploegh and Ir. Martin van Jole at Rabobank for introducing me to the topic and for their continuous support. Their useful feedback, comments, and knowledge supported me throughout the process of writing this thesis. I also wish to thank Rabobank for providing me with all the necessary facilities for the research. My sincere thanks also goes to Dr. Erik Winands, who provided me an opportunity to join his team as intern. Besides my advisors at Rabobank, I am grateful to my supervisor at Utrecht University, Dr. Cristian Spitoni for his insightful comments and encouragement, but also for the hard questions, which inspired me to approach my research from various perspectives. Furthermore, I would like to thank my second supervisor, Prof. Dr. Roberto Fernández

I am grateful to my partner Kay Bogerd who supported me through-out this project and helped proofread this report. Lastly, I thank my parents for their unceasing encouragement and support.

Contents

Acronyms

General Acronyms

AIC Akaike Information Criterion.

BIC Bayesian Information Criterion.

DFAST Dodd-Frank Act Stress Testing.

EBA European Banking Authority.

FFIEC Federal Financial Institutions Examination Council.

KKT Karush-Kuhn-Tucker.

Lasso Least Absolute Shrinkage and Selection Operator.

MAE Mean Absolute Error.

MAPE Mean Absolute Percentage Error.

MLE Maximum Likelihood Estimation.

MSE Mean Square Error.

NBER National Bureau of Economic Research.

NCREIF National Council of Real Estate Investment Fiduciaries.

OLS Ordinary Least Squares.

SWOT Strengths Weaknesses Opportunities & Threats.

Bank Specific Acronyms

AGRI Agriculture.

C&I Commercial & Industrial.

CLD Construction & Land Development.

CO Charge Offs.

CON Consumer.

CRE Commercial Real Estate.

DEP Depository Institutions.

EAD Exposure At Default.

FARM Farmland.

HELOC Home Equity Lines of Credit.

LEASE Leases on Financial Receivables.
LGD Loss Given Default.

MF Multi-Family.

NCO Net Charge Offs.
NFNR Non-farm Non-Residential.
NIM Net Interest Margin.

OTHER Other.

P&L Profit and Loss.
PD Probability of Default.

RC Recoveries.
RES Closed-End Residential Real Estate.
RRE Residential Real Estate.

TL Total Loans.

Macro Economic Acronyms

CPI Consumer Price Index.
CPPI Commercial Property Price Index.

DI Disposable Income.
DJIA Dow Jones Industrial Average.

GDP Gross Domestic Product.

HPI House Price Index.

NROU Natural Rate of Unemployment.

UR Unemployment Rate.

VIX Chicago Board Options Exchange Market Volatility Index.

List of Figures

List of Tables

Chapter 1

Introduction

Setting The continued recurrence of financial crises in the economic history around the world is astounding, especially since most of them seem to follow a similar pattern. Typically, debt and risk gradually accumulate until a sudden panic starts on the markets and a recession ensues. What is even more surprising is that financial authorities have had little power in preventing these busts or reducing their impact, despite their repetitive nature. The most recent economic crisis led to outrage in society concerning the ineffectiveness of risk management and the incapability of banks to protect themselves against adverse events. A central theme was the inability of regulators and risk managers to predict the economic crisis.

Economic crises have a high cost because of their long-lasting negative impact, not only on the economy but also on people's pensions and livelihoods. Therefore, it is important to learn from past mistakes. The complexity of markets and the macro-economic system, currently make it impossible to predict what exactly triggers the occurrence of a crisis or how it will propagate through the economic system. But generally we are able to identify the risk factors driving the boom-bust behavior of the economy. This allows us to construct sensible what-if scenarios in a so-called stress test, which has become an indispensable tool in risk management. It can be described as an assessment of the impact that adverse macro-economic scenarios have on the portfolio of a bank or on general financial stability.

For stress testing we need to model the link between macro-economic developments and banking variables in order to determine the impact of extreme scenarios on bank profitability. The most frequently used approach is bottom-up, where predictions about future profits or losses are made on the most disaggregated level of the portfolio. Because bottom-up methods are generally data-intensive and make it difficult to identify the exact drivers of losses, top-down methods that make predictions on an aggregated level of the portfolio can complement this approach. The advantage of using a top-down model is that it is quick, flexible, and the availability of highly granular data is not essential. Furthermore, parsimonious models can be used to gain insight in the propagation mechanism of macro-economic shocks to profits and losses and can give longterm forecasts. This makes top-down modeling a promising and attractive direction for the development of new stress testing methods.

Contribution Until now, it is unclear whether it is even possible to capture the macro-link in a parsimonious top-down model that makes accurate and reliable predictions. Neither does there exist a consensus on the ideal model specification. In this report we show that precise predictions conditional on a macro-economic scenario can be obtained using a completely data-driven approach to construct parsimonious top-down models for credit loss rates of individual banks. This method can even be used to produce accurate predictions for forecasting horizons of more than five years.

The ultimate goals of this research project are closely related to the desire of Rabobank to further improve its own top-down stress testing framework. The main contribution of this thesis is twofold. First of all, we identify potential pitfalls and limitations related to top-down modeling of credit loss rates. Secondly, the techniques and methods that we employ, can aid in the model selection process. We show that it is possible to use automated feature selection methods to discover the most relevant predictors. In this report we pay special attention to the minimal data requirements and the required granularity of the training data.

Currently, there exists little research that considers the effects of aggregation on the prediction accuracy of top-down models. We find that more granular loan data greatly improves the predictive strength of our top-down stress testing method, since the specificity of macro-economic developments and the composition of the loan portfolio can be taken into account. The results for different loan categories in the portfolio strongly indicate that at least one crisis scenario in the training data is necessary to accurately capture the macro-link. This is not surprising, since data-driven top-down models can only learn from historical events present in the data. Or to put it differently, a data-driven top-down model can only predict the response for stress scenarios that are similar to historical data.

By approaching the top-down model from a statistical learning perspective, we introduce, implement, and test a completely novel approach for model selection of credit loss rates. This is done by regarding the model as a description of the learned relationships in the historical data. We use Adaptive Lasso as an automated feature selection procedure to generate hypotheses about the link between macro-economic developments and credit loss rates for specific loan segments. Such an hypothesis is in the form of a linear model specification. What is remarkable about Lasso methods is their ability to produce *sparse* estimates for coefficients in a linear model. This means that coefficients of non-relevant predictors are shrunk to zero. The technique can even be used when there are more predictors than observations, making it a method that can select the most relevant predictors from a large set of candidates. A favorable attribute of the Adaptive Lasso method is that it produces models that can be assessed by expert judgment. We evaluate the selected models from an economical perspective and use them to predict future credit loss rates conditional on a macro-economic scenario. Our tests show that the predictive strength of the model selected by the Adaptive Lasso method outperforms the currently used methods.

The final part of this report is devoted to a discussion of methods that can reduce the potentially stringent conditions on the minimum amount of training data or alternatively, the limitations on the stress scenario. If the requirement for completely data-driven methods is relaxed, expert judgment can be used to supplement the training data. The advantage of the new machine-learning perspective on macro-economic modeling of credit loss rates is that it gives a suitable framework to learn from past events in a data-driven manner. In the future, this framework could also allow us to incorporate expert judgment and learn from events in other economic systems.

Structure The remainder of this report is structured as follows. The thesis is divided in the parts Background, Data, Model, Results, and Conclusion & Discussion. The Background starts off with Chapter 2, which gives introductions to stress testing, macro-economics, capital planing, and statistical learning. The aim is to familiarize the reader with important concepts and terminology of this thesis. Chapter 3 continues with definitions of the research problems and the formulation of precise research questions. In Chapter 4, we give a concise overview of current advances in the literature regarding top-down stress testing and statistical learning for financial applications.

The Data part consists of two chapters which discuss the data that was used for this research project. For the United States there exists a large publicly available data-set containing structured information on financial statements of commercial banks. Chapter 5 introduces and describes macro-economic and bank-data from the United States and discusses the structure and availability of key banking variables. However, the available bank-data is intended for regulatory purposes. In Chapter 6 we discuss how the data-set can be filtered and adjusted to be suitable for econometric modeling.

The Model part starts with an introduction to Adaptive Lasso in Chapter 7. We derive theoretical results that show that Adaptive Lasso is approximately unbiased, gives sparse and unique solutions, and has desirable convergence properties. Chapter 8 describes how this technique can be used to select models for credit loss rates. The part concludes with a description of our evaluation methods in Chapter 9.

In the Results part we present the empirical results for the models selected by Adaptive Lasso and benchmark models. Chapter 10 deals with the prediction accuracy of averaged loan categories, Chapter 11 shows the results for loan categories taking individual effects into account, and in Chapter 12 we discuss the results for the complete portfolio of individual banks.

In the Conclusion & Discussion part we combine the results to answer the initially posed research questions. The reliability of the model, potential improvements on the method, and future recommendations are discussed in Chapter 14. Subsequently, we address possibilities for the extension of the domain by new learning methods in Chapter 15.

We devoted part of the research project to a literature study, and relevant background information on time-series and regression techniques can be found in Appendices A and B, respectively. For the development of the theory behind the Lasso method, background knowledge in convex optimization is required. An overview of some key results is given in Appendix C.

Part I

Background

Chapter 2

Research Setting

The purpose of this chapter is to provide relevant background information about the setting in which top-down stress testing for banks is used. To that end, we start off with a discussion of stress testing and its link to capital planning. To motivate the importance of modeling credit losses, we break down the impact of several types of profits and losses on the total profitability of a bank during a stressed scenario. To complete the survey of our research setting, we include a concise description of macro-economic modeling and associated challenges. A second objective of this chapter is to introduce and motivate our modeling approach. We give an introduction to statistical learning, and pay special attention to its application to macro-economic modeling. We explain shortly why techniques from this particular field are useful in the context of top-down stress testing.

2.1 Introduction to Top-Down Stress Testing

In this section we present a broad description of current stress testing methods. Two commonly used approaches to the implementation of stress tests are bottom-up and top-down methods. A Strengths Weaknesses Opportunities & Threats (SWOT) analysis is a structured planning method to assess the strengths, weaknesses, opportunities and threats involved in a future project. We engage such an analysis to evaluate the advantages and disadvantages of top down stress testing and motivate why it is a useful tool.

Description A stress test is an assessment of the impact of an extreme but plausible shock to the macro economy on the financial position of the banking sector as a whole or on the condition of an individual bank. Since the global financial turmoil of 2007-2009, stress testing has rapidly grown to be a highly valued tool in risk management [21]. From the perspective of financial authorities such as European

Banking Authority (EBA) or the Federal Reserve, stress testing is a method that can be used to decrease the likelihood and impact of a crisis by testing the financial system's resilience to (global) macro-economic shocks. On the other hand, from the perspective of an individual institution, stress testing can be useful in planning processes by allowing specific adverse future events to be taken into account.

Stress scenarios A stress test starts with the selection of a suitable macro-economic scenario which represents a relevant shock. It is typically described in terms of leading macro-economic indicators such as the unemployment rate, gross domestic product, and price indices. Generally, financial authorities construct macro-economic scenarios that affect the banking sector as a whole. These types of scenarios are driven by systemic risk factors. The effect of such a scenario, for instance a sovereign default, a decline in house prices, or a sharp decrease in oil prices, is then translated to a future time-series of selected macro-economic variables. Banks are instructed to assess the impact of such a scenario on their capital position, Profit and Loss (P&L) statement, and liquidity. The results are then used to assess the quality of the portfolios of individual banks and the resilience of the financial system to the shock.

Another type of risk factors are idiosyncratic risks. These represent those risks that are specific to an individual bank. As such, it is assumed that there is no correlation between the systemic and idiosyncratic risk factors. Examples are a rogue trader causing significant losses in the trading book of a bank, or a failure of the IT systems. Individual banks can also construct their own stress scenarios. Typically, this is done by performing a so-called sensitivity analysis on the portfolio. This boils down to determining which exposures of an institution have the highest risk and applying a negative shock to those exposures. Generally speaking, a suitable scenario needs to capture market-wide, bank-specific or idiosyncratic events that could have a negative impact on the bank's financial position[1] [42].

For this research project we consider longterm systemic risks caused by macro-economic busts. In particular, we focus on the type of scenarios that are prescribed by the annual Federal Reserve stress testing exercise. Such a scenario is expressed in terms of financial indicators such as the Gross Domestic Product (GDP), House Price Index (HPI), or the interest rates. A detailed discussion of these scenarios is provided in Chapter 5.

[1] Actually, there exists an alternative to scenario testing, namely reverse stress testing. In that case, instead of evaluating the influence of specific circumstances, one takes the outcome of a business failure as a given, and identifies circumstances under which this event may occur.

Bottom-Up and Top-Down Historical data can be used to estimate the relationship between macro-economic scenarios and banking variables. Such a model can then be used to obtain predictions of the response of these banking variables to shocks in macro-economic variables. This is a typical approach to stress testing. As was mentioned at the beginning of this section, there exist two distinct approaches to scenario testing; bottom-up and top-down stress testing. In the bottom-up case, the portfolio of the bank is completely disaggregated into sets of similar products. The bank uses its internal models to apply the macro economic shock to each product type. The predicted profits and losses are then aggregated to evaluate the capital position and liquidity of the bank in the macro-scenario.

An alternative approach is the top-down stress testing methodology. For such a method one uses historic time series of the portfolio on an aggregated level, to estimate the relation between macro-economic variables and bank profitability. Consequently, the granularity of the data is much lower. The estimated relation between the macro-economic and aggregated banking variables is then used to predict future losses.

Top-Down (SWOT) analysis The advantages and disadvantages of top-down stress testing can be analyzed by employing a Strengths Weaknesses Opportunities & Threats (SWOT) analysis. Below, we summarize the strengths, weaknesses, opportunities, and threats that are associated with top down stress testing.

- **Strengths**: Since top-down stress testing methods use data on an aggregated level, the data requirements are substantially smaller than those for bottom-up stress testing, which uses more granular data. Also, noisy data may have an adverse effect on model estimation and aggregated data is generally less noisy than granular data. Finally, since estimation is done at a high level of aggregation, the amount of models that need to be maintained is limited compared to the bottom up case. Therefore top-down tests are relatively inexpensive in terms of labor-cost.
- **Weaknesses**: By analyzing an *aggregated* portfolio we lose the ability to capture the sensitivity of the portfolio to specific shocks. A more detailed evaluation of a shock to the loan portfolio, assesses the sensitivity of many product types to a macro economic shock and takes specific exposures into account.
- **Opportunities**: An ideal top-down stress testing method provides a parsimonious description of the relation between an aggregate portfolio and the variables in a stress scenario. A parsimonious model is easier to interpret than a complex model. Such a sparse model may give the opportunity to analyze the results and act upon them. Therefore, a top-down model may contribute significantly to a broader understanding of how the resilience of an institution to a negative

shock can be improved. Simpler models can be easier to generalize, and it might allow us to use a top-down model for forecasts up to five years.

- **Threats**: A concentration of risk exposures in the portfolio can be detected by means of a stress test. If the method is not granular enough, the concentration of exposures on lower levels of aggregation can be overlooked, leading to less precise or inaccurate results. The aggregate predictions represent a broad range of possible outcomes on a more disaggregated level. Since this may confound important implications of the stressed scenario, this can lead to overly optimistic conclusions. Another threat is that we inadvertently use an estimated model for extrapolation. In this case, a linear model may seem appropriate on the training data, whereas the relation between the variable is non-linear in reality. Especially for extreme shocks, a misspecified model can then leave us with substantially misguided information.

To supplement the bottom-up stress tests that are already frequently in use, this research project was initiated to assess the feasibility of the top-down approach and to develop methods for its enhancement. From the analysis above we conclude that, due to the loss of specificity in top-down methods, it is crucial to determine an appropriate aggregation level of the portfolio, which adequately captures the impact of the specified scenario. More explicitly, we want to determine an aggregation level of a bank's loan portfolio, such that it is suitable to estimate a model that forecasts credit losses conditional on paths of exogenous macro-economic variables.

The price or loss processes of different products may be correlated in a way that cannot be explained by a common dependency on the macro-economic variables. Such correlations can be difficult to estimate, but for a top-down method these do not have to be taken into account directly. Moreover, since the data requirements are expected to be substantially smaller, a top-down test can be executed quicker and it is easier to adapt it to changing circumstances. Lastly, top-down models could produce accurate predictions further into the future than bottom-up methods. Therefore, a top-down stress testing method is a great complement to bottom-up procedures.

Stress Testing and Forecasting Forecasting is the process of analyzing trends on past data to predict future events. A model used for a stress test is an example of a forecasting model since it predicts future losses. On the other hand, it is assumed that the future macro-economic developments, as specified by the stress scenario, are known and hence the results are only predictions conditional on these scenarios.

The main advantage of stress testing is that it provides an assessment of a bank's financial position in adverse scenarios whose occurrence is not predicted or captured by existing methods. In fact, economic developments are not just uncertain, we do not even know the range of possible scenarios that can occur. Therefore it is extremely challenging, if not impossible, to predict such extreme but plausible events beforehand[2]. However, a stress test allows us to assess the impact of adverse events on the portfolio. As such, stress testing can be used to protect institutions and the financial system against macro-economic shocks.

Macro-economic modeling has been a topic of much debate, since it is subject to the so-called *Lucas critique*. It can be argued that it is naive to predict the effect of changes in the economy based on historical relationships between highly aggregated data. One can only predict future states by assuming that all circumstances remain equal (i.e. the world does not change). However, for stress testing the Lucas critique is a problem of lesser degree, because one of the assumptions of a stress test is that the future is already (partially) known and all other relevant circumstances remain equal. Caution should however be taken with the use of historical data for estimation, which may not be representative of the current economic conditions and regulations. Stress testing should therefore be understood as a *complement* to the existing risk framework.

2.2 Macro Economics and Capital Planning

To gain a better understanding of the economic setting of the research project we studied generally accepted theory regarding credit, debt, and macro-economics. Although a thorough economic interpretation of fluctuations in bank profitability reaches beyond the scope of this thesis, we give a succinct introduction.

Macro-economics deals with the structure, performance, and behavior of the economy as a whole. A key concept is the so-called business cycle, which describes the phenomenon that aggregate economic activity tends to fluctuate over a certain period of time. Typically, a business cycle consists of the prosperity, recession, depression, and recovery phase. These phases can be characterized by economic measures such as the unemployment rate, GDP growth, and interest rates.

Many of the underlying processes that have been identified by economists are said to be of a pro-cyclical nature, meaning that they generate a positive feedback sys-

[2] In the 1920s Knight and Keynes theorized about the different nature of risk and uncertainty. In 1989 Bausor formulated the concept as follows: "Sample spaces must contain all possible future outcomes, including the "true" outcome, and this inclusivity must be known. No possibility can be neglected, overlooked or unimagined. States of the world, however, are not ontologically existential. Through effort and skill, they must be conjured up from the imagination, and imagination is always vulnerable to fallibility. People constantly experience previously unimagined phenomena, and the potential for surprise remains ubiquitous. Since sets of imagined possible outcomes cannot be known to be complete, no standard for measuring the relative strength of beliefs exists, and distributing weights so that their sum equals one-the construction of a probability measure-becomes invalid and meaningless."

tem. The behavior of these fluctuations in economic activity is generally unpredictable since typically, economic contractions and expansions occur aperiodically, have asymmetric impact, and are of variable duration [55]. The banking system has an important role in the propagation of this cyclical behavior.

Traditionally, a bank was a place where commodities, money or coin could be stored safely. This practice dates all the way back to ancient Mesopotamia where depositors collected a token in exchange for their commodities which could be traded with third parties and on which interest could be charged [50]. Today, commercial banks play a central role in our economy. They regulate (electronic) money transactions, offer insurances, manage investments, provide currency exchange services, price complex financial products and so on. However, the core business of banking still is its role as an intermediary between lenders and borrowers.

Banks play a crucial role in the macro-economy by supplying borrowers with credit. They operate by transforming deposits into loans with different characteristics such as size, maturity, and currency. They take margins on these transformations and charge fees to turn a profit [12]. Arguably, the availability of credit and the amount of debt is closely related to longterm macro-economic developments [55].

In its capacity as an intermediary, the bank manages its assets, liabilities and equity. The ratios between assets, liabilities and equity need to comply with the regulations of central banks and other financial authorities whilst generating profit for the bank and its potential shareholders. Profits are generated by maintaining a margin on the interest rate paid by debtors and collected from creditors and by charging fees for other financial services. Losses are incurred by lower valuation of (trading) assets and defaults on loans [12]. Default rates and asset values are susceptible to macro-economic shocks.

In order to inform shareholders and regulators about the condition of the bank and its profit results, most commercial banks regularly publish financial statements such as the balance sheet and the Profit and Loss (P&L) statement. The latter is a summary of the revenues, costs and expenses incurred during a certain period of time. It typically contains items such as interest income, interest expense, non-interest income, non-interest expense, operational costs, provisions, and taxes. Each of these items can be further disaggregated into subcategories. Below we give a brief description of the items that can typically be found on the P&L statement.

- Net Interest Income: Banks generate income by managing the spread between interest rates paid on deposits and interest rates charged on loans. As an intermediary, a bank transforms deposits into loans which typically have a longer maturity than the deposits. This mismatch between assets and liabilities causes interest rate risk to a bank, since a discrepancy between the interest charged and received changes. Therefore bank profitability is sensitive to systemic shocks to the interest rates.

- Non-Interest Income: Non-interest income is important for banks to insulate themselves against interest rate risks. This item consists of fees and commissions which can be charged on brokerage services, insurance services, etc. Another component of non-interest income are profits generated by trading activities and gains from selling securities. Generally, the income from fees and commissions is relatively stable and trading income is viewed as highly volatile and unpredictable. The main risk factors for non-interest income are both systemic and idiosyncratic.
- Provisions: Reserves for loan and lease losses that are kept on the balance sheet are depleted as credit losses are incurred. These loss reserves are replenished by provisions, which are deducted from the bank's revenues. Provisions are subject to the risk that the underlying securities of loans decreases in value and to the risk that the borrower defaults. Provisions are affected by systemic shocks to the macro-economy.
- Operational Costs: A bank is a business and therefore it has operational costs, which includes staff costs, administrative costs, depreciation and amortization. When an exogenous systemic negative shock to the entire banking sector occurs, the regulations in the country that the bank operates in are important. A flexible market allows a bank to efficiently reduce its labor costs and thus increase its profitability. Risk factors include both systemic and idiosyncratic risk.
- Taxes: Finally, taxes must be deducted from the profits. The amount of taxes is determined by (inter)national laws and regulations.

In order to get a sense of the typical size of some items on the P&L statement, we show some of the annual profits and losses for Rabobank between 2004 and 2014 in Figure 2.1. It can be seen that in the last decade the items net interest income, non interest income, and credit losses have all been sufficiently large to significantly impact the total profitability of Rabobank.

For this research project we decided to focus on predicting credit losses conditional on macro scenarios. Since these are typically sensitive to long-term macro-economic developments and shocks, it is an interesting item on the P&L statement. Further research is necessary to develop methods for the stress testing of other items on the P&L statement.

The capital position of a bank must be in compliance with rules set by financial regulators. The change in an institution's capital over a period of time depends on the profits (losses) incurred during that period. Therefore, the forecasting of items on the P&L statement conditional on macro-economic developments is meaningful

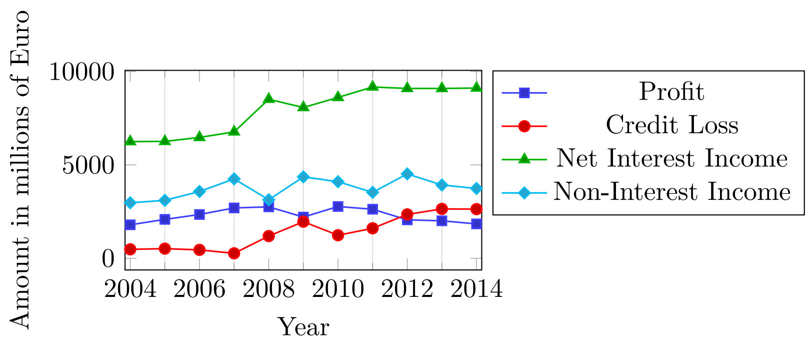


Figure 2.1: Rabobank's Profit & Loss Statement

for a stress test on the capital position of the bank. The item provisions appears directly on the income statement as an expense. It is used to replenish the allowance for loan and lease losses that is kept on the balance sheet. The amount of provisions that is booked is calculated according to a scheme that takes into account the credit losses in the preceding period, and the expected losses in the following period. We assume from here on that provisions can be calculated directly from credit losses, and focus solely on the modeling of credit loss rates [47].

The call for a sound capital planning process is a serious aspect of the Basel III accord. This condition specifically requires a bank to assess the risks to which it is exposed and to consider the potential impact on earnings and capital caused by an economic downturn. According to Basel III, a robust stress testing framework conservatively captures the change in risk factors in forward-looking scenarios. The impact of a stress test should then reflect changes in income, loss, exposure, risk-weighted assets and a change in capital needs [42]. To comply with these directions, a diverse and repeatable stress testing framework needs to be incorporated into banks' capital planning processes.

Rabobank wishes to adopt such a forward looking approach to capital planning. At Rabobank, a capital planning tool is currently being extended, which should ideally include a module for top-down stress testing. This enables Rabobank to quantitatively consider the potential impact of future events on its earnings and its capital position. The main goal is to develop new methods than can predict losses conditional on macro-economic scenarios up to five years into the future. The methods for credit losses that were developed during our research project, can be used in the development of such a broad framework that focuses on the assessment of the capital position of Rabobank, given an adverse future scenario for the macro-economy.

2.3 Statistical Learning and the Lasso

Machine learning is a field of computer science that studies algorithms that can be used to learn from data and make data-driven predictions. Originally, machine learning was defined as 'a field of study that gives computers the ability to learn without explicitly being programmed'. Its aim is to build a model from example inputs. The combined field of machine learning and statistics is often referred to as statistical learning. With the rapid expansion of problems concerning large datasets, statistical learning techniques have become exceedingly popular in a variety of fields such as biology, business, and medicine.

Typically, statistical learning methods can be classified as either supervised or unsupervised. Broadly speaking, supervised statistical learning tools are used to build a model for estimating or predicting an output based on a set of inputs. As such, supervised statistical learning is the process of extracting regularities from large and complex data sets to learn about the environment and make predictions. When the data is unsupervised, there is no output data, but relations and structures can still be learned from such a data set. A top-down model that relates banks' credit losses to macro-economic scenarios surely falls into the supervised learning category.

So when can we say that a computer program is learning? A common definition is that 'it learns from experience E with respect to task T and performance measure P, if its performance at tasks in T as measured by P, improves with experience E'. In a typical statistical learning problem response variables $y \in \mathcal{Y} \subseteq \mathbb{R}^n$ have to be predicted from a set of feature variables $x = (x_1, \ldots, x_n) \in \mathcal{X} \subseteq \mathbb{R}^{n \times p}$. We say that the labeled data $(y_1, x_1), \ldots, (y_n, x_n) \in \mathcal{Y} \times \mathcal{X}$ are distributed according to a density $p(x, y)$ and we assume that this distribution never changes. The idea is that the computer is presented with input data $x \in \mathcal{X}$ and output data $y \in \mathcal{Y}$, and automatically learns a mapping $f(x) \to y$. In this case the learning task is the mapping f and the experience is the number of labeled data. A possible performance measure is the sum of squared residuals

$$\sum_{i=1}^{n} \left(f(x_i) - y_i\right)^2 .$$

When f is a linear mapping, this performance measure corresponds to the well-known Ordinary Least Squares (OLS) regression. The disadvantage of OLS is that the structural form of the model needs to be specified in advance. A renowned supervised learning algorithm that does not need a pre-specified model, is the Least Absolute Shrinkage and Selection Operator (Lasso) method, which was introduced by R. Tibshirani in 1996 [2]. It minimizes the sum of squared residuals, but adds a penalty based on the absolute size of the coefficients of the linear model. In Chapter

7 we shall proof that this methods sets coefficients to zero, resulting in a sparse and parsimonious model.

The idea of the Lasso method is to minimize,

$$\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{2.1}$$

where λ is a tuning parameter that controls the complexity of the resulting model. If $\lambda = 0$, the problem reduces to OLS regression, and when $\lambda = \infty$ all of the coefficients will be set to zero. The key to a good solution is that there exists an optimal λ, which balances the variance and bias of the model, resulting in low prediction errors.

This much acclaimed procedure possesses many favorable statistical properties[3], and allows for data-driven model discovery. Another fortunate attribute of the method is that it can, unlike classical methods, estimate the coefficients β of a linear model, when the number of predictors far exceeds the number of observations. It can be shown that these predictions are unique, under very light conditions. Also, feature selection for a statistical model can be fully automated using Lasso methods. A slightly more involved version of Lasso is Adaptive Lasso, which can be shown to give approximately unbiased estimates of β. Moreover, it possesses attractive convergence properties. During the research project we chose an initial formulation by determining potentially relevant predictors. We developed techniques based on Adaptive Lasso to effectively model and predict credit loss rates in stressed scenarios for banks.

The discussion of macro-economics and bank profitability in Section 2.2 suggests that macro-economic data and credit loss rates are related in a complex way. In fact, as Doornik and Hendry (2015) put it, we are dealing with inaccurate measurements of non-stationary and high-dimensional systems that are correlated, probably non-linear and on top of that evolving in time. This makes model selection a highly challenging exercise. For such complex systems all important predictors, their interactions and non-linear dependencies need to be taken into account. Since the correct model specification cannot be known in advance, the substantially relevant effects need to be discovered in a data-driven way. Omitting relevant variables may lead to erroneous conclusions, while over fitting a model can lead to spurious regression. Therefore, we risk selecting a significantly misspecified model [51]. However, a sound statistical analysis requires a model which is adequately specified. To advance the model selection problem, we propose the use of Lasso methods to discover a parsimonious model.

Machine learning is a fast-growing area of research, which has received much praise,

[3] The finite-sample and asymptotic properties of Lasso will be derived in Chapter 7

but also a lot of criticism. Most of it is aimed at the large amount of false positives that are detected in data sets [46]. In this case, accuracy is overrated by ignoring the non-existing relations that were found by the procedure as well. A second aspect of the critique is that machine learning algorithms give no insight in causality or underlying mechanisms and hence can only give associations. This can be harmful since a researcher that cannot explain which causes what, cannot be sure what may cause the correlations in the model to break down. This has led researchers in some fields to the conclusion that machine learning 'is not a big deal, just another tool'[4]. What is meant by this, is that the results of machine-learning methods should not be trusted blindly and need to be evaluated from a theoretical point of view.

For high-dimensional data problems with many predictors, machine learning methods are an important part of the toolbox. Lasso methods are highly valuable tools because the discovered models can easily be evaluated from a theoretical point of view. High-dimensional data problems can, for instance be encountered in macroeconomics. A major problem is that when statistical results cannot be validated by scientific experiments, we must rely solely on statistical inference. Therefore, statistical learning methods need to be equipped with measures of uncertainty, reliability and significance [25]. Another advantage of Lasso estimators is that some of these measures can be derived, and much work is being put forward to improve these measures. Importantly, for a top-down model the selected predictors from the initially chosen set of candidates can be evaluated from an economic point of view.

[4] This was one of the conclusions of the panel discussion Large Data and Hypothesis-Driven Science in the Era of Post-Genomic Biology with Bruce A. Beutler and Nobel-laureates J. Michael Bishop, Moderator Stefan H.E. Kaufmann, Brian Schmidt and Jules A. Hofmann in Lindau on July 2, 2014.

Chapter 3

Research Questions

In the previous chapter we explained that the focus of this thesis is on the modeling of credit losses for top-down stress testing purposes. We develop a top-down model to assess the extend to which it can be used for the stress testing of credit losses. In this chapter we formulate the research questions that we aim to resolve in this report, and the main goals of our research project. To that end, we introduce relevant notation and provide a description of the problem setting. We proceed with precise statements concerning the research problems. We conclude by establishing our research questions.

3.1 Assumptions and Definitions

From here on, we let time be indexed by $t = 1, 2, \ldots, n_T$ at intervals of one quarter of a year. The loan portfolio consists of many individual loans, which are organized in *loan categories*. Loss rates on such categories are assumed to be uncorrelated, apart from a common dependency on macro-economic developments and bank-specific predictors.

Definition 3.1. (Loan Categories) The individual loans in a loan portfolio $\mathcal{P}$ are aggregated in loan categories P_i with $i = 1, \ldots, n_C$, based on similarity in their characteristics (such as maturity or type of underlying security). It holds that $P_j \cap P_k = \emptyset, \forall\, j \neq k \in \{1, \ldots, n_C\}$.

In the general framework for bottom-up stress testing models, absolute credit losses have three components. Firstly, a loss is only suffered when the borrower defaults. Secondly, the maximal loss that the bank suffers is the exposure it has towards the borrower. And finally, the amount that is lost in case of a default depends on the value of the underlying security at that time. For instance, when a customer receives a mortgage, the value of the house serves as a security to the bank. If the customer were to default, then part of the current value of the house R can still be recovered. We use the well-known terms Probability of Default (PD), Exposure At Default (EAD) and Loss Given Default (LGD), respectively, to describe the three components of credit loss.

Since the size of loan portfolios changes over time and differs between institutions, losses need to be made comparable to model them effectively. A natural way to accomplish this is to focus on credit loss (CL) rates instead of absolute losses. PD, LGD, and EAD are related to credit losses in the following way,

$$\frac{\mathrm{CL}}{\mathrm{EAD}} \sim \mathrm{PD} \times \mathrm{LGD} = \mathrm{PD} \times (1 - \mathrm{R}). \tag{3.1}$$

On the other hand we have Net Charge Offs (NCO) rates, which can be interpreted as a measure for the loss rate on each dollar that is lent out. We let $\mathrm{NCO}_{i,t}$ denote the net charge offs for one loan category P_i at time t. Charge Offs (CO) are the amount of loans that is written off due to a default, and Recoveries (RC) are the amount of the defaulted loans that is recovered. Lastly, Total Loans (TL) is the total amount of outstanding loans.

Definition 3.2. (Net Charge Off Rates) NCO rates on a loan category $P_i \in \mathcal{P}$ are a function of Charge Offs (CO), Recoveries (RC), and Total Loans (TL). It can be calculated as follows,

$$\mathrm{NCO}_{i,t} = \frac{\mathrm{CO}_{i,t} - \mathrm{RC}_{i,t}}{\mathrm{TL}_{i,t}}. \tag{3.2}$$

From 3.2 it follows that we can only obtain NCO rates for discrete time intervals since we need to estimate this rate by charge offs and recoveries that are accumulated over time. To illustrate this we present the average of the accumulative CO, RC, and their difference for all commercial banks in the United States in Figure 3.1. The NCO rate can be obtained by dividing the difference between the charge offs and recoveries by the total outstanding amount of loans.

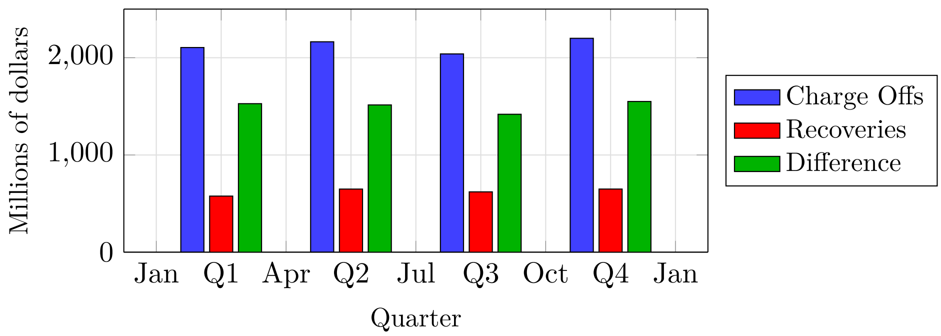


Figure 3.1: Average quarterly charge offs and recoveries for all US banks in 2014.

Alternatively, we can interpret $\text{NCO}_{i,t}$ as a continuous process representing the annual net percent loss at time t on the total of outstanding loans within loan category $P_i \in \mathcal{P}$. But for this research project, we define $\text{NCO}_{i,t}$ as the average NCO rate in the time interval $(t-1, t]$, and calculate the rate in each quarter t by using accumulated charge offs and recoveries in equation 3.2. In Chapter 6 we discuss possible issues with obtaining discrete-time estimates for the NCO rate from reported Charge Offs, Recoveries, and Total Loans regulatory data.

Where bottom-up stress testing methods typically estimate PD,LGD and their correlations to model losses for individual loans, we suggest to model Net Charge Offs (NCO) rates for an entire loan category P_i directly. First we note that equation 3.2, can be rewritten as,

$$\text{NCO}_{i,t} = \frac{\text{CO}_{i,t}}{\text{TL}_{i,t}} \left(1 - \frac{\text{RC}_{i,t}}{\text{CO}_{i,t}}\right). \tag{3.3}$$

The relative amount of charge offs compared to the total outstanding loans can be interpreted as a default rate d_i for each dollar. Similarly, we can define r_i as a recovery rate for each outstanding dollar in loan category P_i. An intuitive formulation of equation 3.3 is therefore,

$$\text{NCO}_{i,t} = d_i(1 - r_i). \tag{3.4}$$

The NCO rates as defined in equation 3.2 can thus be interpreted as an aggregated, top-down version of the credit loss rates in equation 3.1[1]. In Chapter 4 we discuss common practices with respect to top-down stress testing, and it turns out that in current literature it is common to model Net Charge Offs (NCO) rates instead of absolute credit losses in the top-down setting.

Typically, net charge off rates are presented as a percentage. Theoretically, the rate should be between 0% and 100%, because the bank cannot lose more than the exposure it has towards the borrower and the recovered amount should not exceed this exposure. Due to the processes within the bank that are involved with writing off debt and recovering loans, rates below 0% and above 100% do occur in practice.

The goal of top-down stress testing is to make forward predictions for NCO rates, conditional on macro-economic scenarios, keeping all other circumstances constant. The forecasting or *stress* horizon is the period for which these predictions are made.

Definition 3.3. (Stress horizon) The stress testing horizon h is the number of quarters for which a forecast of the NCO rate conditional on a stress scenario is given.

Typical values for the stress horizon are in the range of 4 to 20 quarters. Following

[1] For a comprehensive example of this principle see Wang (2013) [41]

stress testing practices in the United States, we will focus on forecast horizons of 9 quarters in particular. The optionality to forecast even further ahead is one of the most attractive features of top-down compared to bottom-up stress testing. For Rabobank this is an important consideration and therefore we also take long-term forecast of up to 5 years into account.

Definition 3.4. (Macro Stress Scenarios) Let M_t denote a $(p \times 1)$-vector of the values at time t of macro economic variables that are included in the stress scenario. The scenario $M_{t,t+h} = (M_t, M_{t+1}, \ldots, M_{t+h})$ is a discrete description at regular (quarterly) intervals of the future state of the macro-economy in an adverse event.

For the prediction of loan loss rates for individual banks, specific characteristics of the bank in question may also be relevant in explaining variation in NCO rates for individual banks. Furthermore, the use of NCO rate data for multiple banks may help to separate losses due to idiosyncratic risk factors from those due to systemic macro-economic developments. Estimating individual effects might help us control for factors that are independent of macro-economic variables. To that end, we denote banks by $b_j : j = 1, \ldots, n_B$ and let $B_{j,t}$ be a $(q \times 1)$-vector of (relevant) bank-characteristics of bank b_j at time t. From here on, we let $\text{NCO}_{i,j,t}$ denote the NCO rate at time t for bank b_j and loan category P_i.

In summary, we want to make accurate predictions for future net charge off rates $\text{NCO}_{i,j,t}, \ldots \text{NCO}_{i,j,t+h}$ on loan categories $P_i \in \mathcal{P}$, for bank b_j, given a macro economic scenario $M_{t,t+h}$, until forecasting horizon h. A very general form of a top-down model could be,

$$\text{NCO}_{i,j,t+h} \sim f_{i,j}(M_{t,t+h}, B_t, \text{NCO}_{i,j,t}) + \epsilon_{i,j,t+h}, \tag{3.5}$$

where $f_{i,j}$ is a function of unknown form for loan category P_i, and $\epsilon_{i,j,t}$ is an error term with unknown structural form. Note that there are no future paths for bank characteristic variables given in stress scenarios and therefore $f_{i,j}$ is a function of B_t and not of the path $B_t, \ldots B_{t+h}$.

In a stress test we assume, that all circumstances remain equal over the stress horizon, with the exception of the macro-economic stress variables whose values are known. This simplifies the task of predicting future NCO rates, since no paths need to be generated for other risk factors. For model estimation, we would like to control for these independent risk factors. Since we are interested in the opportunities of stress testing for an individual bank and NCO rates differ among banks, we would like to uncover a relation between bank characteristics and credit losses. These bank-specific variables may be able to explain heterogeneity in NCO data.

We note that, there exists no theoretical models that link changes in the macro-economy to aggregated NCO rates. The macro-economy and the financial system are both extremely large and highly complex systems whose dynamics are not well understood [51, 46]. The challenge is therefore the discovery of a link between macro-economic developments and bank profitability, preferably using a large set of data. To determine whether top-down stress testing is feasible and to which extend it can be used for real applications, we make assumptions about $f_{i,j}$ in equation 3.5 and show that we can construct a model that can produce accurate long-term predictions.

3.1.1 Mathematical Notation

In the remainder of this thesis we use the following notation. The number of predictor variables is p, and the number of observations is denoted by n. We let Y be a $(n \times 1)$ vector of response variables. The design matrix X is an $(n \times p)$-matrix of predictor variables. We let X_i represent the ith row of X. Furthermore, we let $||\cdot||_2$ and $||\cdot||_1$ denote the $\mathcal{L}^2$ and $\mathcal{L}^1$ norm, respectively. The $\mathcal{L}^2$ and $\mathcal{L}^1$ norm of an $(p \times 1)$ vector β can alternatively be denoted by $\sum_{j=1}^{p} \beta_j^2$ and $\sum_{j=1}^{p} |\beta_j|$, respectively. For the derivation of Lasso results in following chapters, we assume that the columns of X are centered and scaled to have mean 0 and variance 1, unless noted otherwise.

3.2 Research Questions

Each of our research questions serves as a stepping stone towards a general method for top down estimation and prediction for stress testing. The resulting top-down model is used to explore the opportunities and assess the reliability of top-down stress testing. We identify several aspects of this exercise, and for each aspect we give a short introduction to the problem and conclude with the formulation of research questions. For the remainder of this section, we assume that an adverse macro-economic stress scenario $M_{t,t+h}$ is given, and unless noted otherwise, we consider a fixed loan category P_i and drop the subscript i from our notation.

3.2.1 Predicting Average Net Charge Off Rates

We consider the average NCO rate for all commercial banks. That is, we calculate,

$$\mathrm{NCO}_t^{\mathrm{av}} = \frac{\sum_{j=1}^{n_B} (\mathrm{CO}_{j,t} - \mathrm{RC}_{j,t})}{\sum_{j=1}^{n_b} \mathrm{TL}_{j,t}}. \tag{3.6}$$

We know that the macro-economy affects banks' revenues and losses, but we do not know the exact relation between macro- and bank-variables such as NCO rates. Where theoretical models are available for the bottom-up case, there exists no theory-based model for top-down models. It is reasonable to expect that the macro-economic effect on NCO rates may be of a nonlinear nature [51]. Because we are interested in prediction of NCO rates in *extreme* but plausible events, we need to consider nonlinear models, since linear approximations typically do not perform well for stress scenarios.

Considering many of such nonlinear transformations and interactions between macro-economic variables quickly leads to a staggering amount of potential predictors. Therefore we want to use Adaptive Lasso to select the substantially relevant transformations. Recall from the previous section that a stress scenario is specified in discrete time until some forecasting horizon h. Let $\mathcal{F}_t$ be the natural filtration of the real-world probability space. We assume that the expected future NCO rates conditional on $M_{t,t+h}$ can be obtained by estimating α, β, and f in the following way,

$$\mathbb{E}[\mathrm{NCO}^{\mathrm{av}}_{i,t+h}|\mathcal{F}_t, M_{t,t+h}] = \alpha + \beta f(M_{t,t+h}), \tag{3.7}$$

where the function f performs linear and non-linear transformations on the macro-economic data and β is a vector of coefficients. Since we do not know beforehand whether the NCO rate process is auto-regressive, we consider both the inclusion and exclusion of auto-regressive error terms. Applying Adaptive Lasso to estimate the model in 3.7, allows us to answer our first set of research questions:

i *Can we use automated feature selection by employing Lasso methods to identify the model specification?*

ii *Can we construct a model that gives accurate predictions for NCO rates conditional on exogenous macro-economic paths?*

iii *Do we need to consider auto-regressive model specifications?*

The macro-economic effects that we want to capture in a top-down model are of a longterm nature. Since we use historic data to learn regularities about the connection between NCO rates and the macro-economy, we require a large amount of data. The predictions of the average model in a stress scenarios can be evaluated. Since we assume that the loan portfolio can be disaggregated into n_C loan categories, we can make predictions for NCO rates $\mathrm{NCO}^{\mathrm{av}}_{i,t}$ for $i = 1, \ldots, n_C$. The range of the relevant historical macro-economic data compared to the range of the macro-economic variables in the stress scenario differs per category. By applying our method to average NCO rates on different loan categories P_i, we shed light on the following research questions:

i *What are the minimal data requirements for top-down modeling?*

ii *How do data limitations affect the prediction accuracy?*

3.2.2 Predicting Net Charge Off Rates with Individual Effects

The main objective of the aggregate estimation exercise is to develop a method (or model) that can be used to discover the relation between credit losses and macro-economic developments. Since we are particularly interested in systemic effects that impact the loan portfolio, it can be convenient to separate idiosyncratic effects that are suffered by individual banks due to the specific circumstances of that bank, from systemic risk drivers. That way, we can control for other influential circumstances and obtain forecasts, where all other conditions remain the same.

We considered the losses for multiple banks at once in the average model. Another option is to use panel data models[2].

A distinction can be made between bank-specific variables that evolve over time such as the capital ratio, and those that are more or less constant over time such as geographical location. We let $B_{j,t}$ be time-dependent bank-specific variables. When we use panel data, we assume that for banks $j = 1, \dots, n_B$,

$$\mathbb{E}[\mathrm{NCO}_{i,j,t+h}|\mathcal{F}_t, M_{t,t+h}] = \alpha_{i,j} + c_{i,j}\left(\beta f(M_{t,t+h}) + \gamma B_{j,t}\right), \tag{3.8}$$

where f is a function that performs non-linear transformations on the macro-economic data. Using this model, we consider the following research question:

i *Can bank-specific variables help explain variability in the credit losses between banks?*

ii *If we assume that $\gamma = 0$, can we then obtain accurate predictions for $NCO_{i,j,t+h}$? Or, is the inclusion of bank-specific effects necessary?*

3.2.3 Forecasting NCO Rates for Individual Banks

From the SWOT analysis of the advantages and disadvantages of top down stress testing in Section 2.1, we concluded that it is prudent to strike some middle ground between bottom-up and top-down methods, in the sense that we select loan categories $P_i \in \mathcal{P}$ which allow us to capture the specificity of a macro economic shock and aggregate the predictions. Moreover, the specific composition of an individual bank's portfolio may play an important role in the development of its NCO rates in a stress scenario. For instance, it is typically so that the losses on credit card loans

[2]In Appendix B we give a concise review of modeling with longitudinal data (also known as panel data).

are much higher than losses suffered on real estate loans. Therefore, a bank with a larger share of credit card loans in its portfolio can have higher loss rates than a bank with a relatively large share of real estate loans in its portfolio.

Essentially, this tells us that it is necessary that we determine a level of aggregation which is suitable for the prediction of the response of credit losses to shocks on selected macro economic variables. After a model for the estimation and prediction of NCO rates of an arbitrary loan category is found, the next step is to disaggregate the total loan portfolio in such a way that sensitivity to shocks and accuracy of the prediction is preserved, whilst keeping the model parsimonious and reducing noise. For the selection of the aggregation level of the top-down model the availability of (publicly) available bank data should be considered. Discovering the ideal aggregation level comes down to finding an optimal combination of top-down and bottom-up methods, as was previously discussed in Section 2.1.

We assume that the loan categories can be organized in a hierarchical structure. We consider three levels of this structure which are referred to as *aggregation* levels A_1, A_2, and A_3, where A_1 is the lowest and A_3 is the highest level in the hierarchy. Each level consists of aggregated loan categories such that for all loan categories $a_1 \in A_1$ we have that $a_1 \subseteq a_2$ for some $a_2 \in A_2$ and for all $a_2 \in A_2$ there exists some $a_3 \in A_3$ such that $a_2 \subseteq a_3$.

For each aggregation level we can predict NCO rates for an individual bank, using the methods developed in the previous two sections, by combining the predictions for different loan categories. This can be done by taking a weighted average of the rates for the different categories,

$$\mathrm{NCO}^{\mathrm{tot}}_{j,t+h}(A_k) = \sum_{a_i \in A_k} \mathrm{NCO}_{a_i,j,t+h} \frac{\mathrm{TL}_{a_i,j,t}}{\mathrm{TL}^{\mathrm{tot}}_{j,t}}, \tag{3.9}$$

for $k = 1, 2, 3$.

This final method gives predictions for individual banks' credit loss rates conditional on a macro-economic scenario. The results of this model allow us to answer the following research questions.

i *What is the influence of the granularity of the NCO data on the accuracy of top-down models?*

ii *What prediction accuracy can be obtained by a top-down stress tests on the loan portfolio of an individual bank?*

iii *What stress horizons are feasible for top-down stress testing?*

3.2.4 Delimitations

The methods described in the previous section lead to a complete top-down model. We use this method to gain a better understanding of the possibilities and limitations of top-down stress testing. The models that we use are not yet directly applicable for Rabobank, because the results of the presented models are determined on data from the United States and the reliability of the method and resulting models needs to be evaluated further. The following topics are only discussed briefly or will be addressed in the discussion part of this report.

Once we have obtained a model to predict the NCO rates for loan categories in the aggregate bank loan portfolio, or individual banks' loan portfolios, what remains is to test its accuracy and reliability. The use of real observations is troublesome, especially for individual banks, since the noise in the observations or the effect of idiosyncratic events may be substantial. Moreover, there is only one sample of the macro-economic path and we cannot test the resulting model by a controlled experiment. This makes the evaluation of any model more challenging.

Part of a lack in reliability of the predictions might be due to the input of a stress scenario that is too dissimilar from the macro-economic developments in the training data. Note that stress testing is not a tool to predict the future, but rather a method that gives expectations of the NCO rates, based on the knowledge of today. Let $\mathcal{D}$ be the range, or domain, of the data on which a model for the prediction of NCO rates is estimated. The domain of a model can be specified as a function of the training data. Assume the training data is given by response variable $y \in \mathbb{R}^{n_T}$ and predictors $x \in \mathbb{R}^{n_T \times p}$. A possible definition of the domain is,

$$\mathcal{D} = \{d \in \mathbb{R}^n | \min_{1 \le i \le n_T} x_{i,j} \le d_j \le \max_{1 \le i \le n_T} x_{i,j}, d_j, \forall j = 1, \dots, p\}. \tag{3.10}$$

For each top-down method a suitable input domain for stress scenarios should be defined. Furthermore, the user needs to be aware of the implications of using more extreme scenarios for prediction in a stress test. This boils down to determining the effect of limiting the range of the estimation data in each dimension. Extrapolating a model poses risks, since the data contains no information about the relationship between macro and bank variables outside of the domain.

It could happen that, especially when there is little available data, that the response of NCO rates can only be reliably predicted for a very limited amount of stress scenarios. Since easily accessible bank data is scarce in Europe and the Netherlands, options to extend the range of macro scenarios for which we can predict NCO rates need to be explored. In particular, it would be interesting to investigate whether we can use functions f^{US} learned on data from the United States to help discover f^{EU}. Another option is the quantitative incorporation of expert judgment.

Because there was a limited amount of data available for the European case, we only considered the prediction of NCO rates on data from banks in the United States. In the discussion part of this thesis we present recommendations based on this research project for the continued improvement of top-down methods for Rabobank.

The above discussion can be summarized by the following remaining questions:

i *How can we measure the accuracy or reliability of a stress test?*

ii *Can we obtain reliable confidence intervals for the predictions of a top-down stress testing model?*

iii *Can the domain of a top-down model be extended?*

3.3 Summary

We contribute to the extension of Rabobank's top-down stress testing framework by determining what can be expected from such an exercise. We model NCO rates, which are closely related to credit losses, in three steps. First we consider averaged NCO rates across banks for different loan categories. Then we use a panel data model to incorporate individual effects and scaling factors. Lastly, the individual NCO rate models for loan categories are aggregated to produce models for the total loan portfolio. In all models we consider linear and non-linear transformations of the macro-economic data.

We employ statistical learning methods to predict future NCO rates, conditional on future developments of the macro-economy in the United States. We focus on stress horizons of nine quarters, but also take long-term stress horizons of five years ahead into account. Since the macro-economic paths are assumed to be known, the task is to discover the relation between NCOs and stressed scenarios. The resulting model is then used to determine a feasible range for the prediction accuracy of top-down models and assess the effects of aggregation and the limitations set by the scope of the training data. We try to capture the heterogeneity in different banks' NCO rates by considering the loan composition, individual effects, and scaling factors. Furthermore, we show that Adaptive Lasso can be used to automatically obtain a parsimonious linear model specification, that gives accurate predictions. The model is compared to linear and auto-regressive benchmark models to show its superior forecasting ability. Part of the thesis discussion is intended to indicate how the method we use to answer the research questions can be exploited by Rabobank.

Chapter 4

Literature

The literature on top-down stress testing credit losses is relatively scarce. However, the recent global financial crisis, after which the Federal Reserve initiated obligatory stress tests for large financial institutions, sparked a renewed interest in top-down stress testing and related methods. Since nobody saw the financial turmoil coming, more attention was focused on methods that help to protect the financial system against severely adverse events.

We give a detailed overview of some existing techniques for top-down stress testing credit losses[1]. In this chapter we discuss the methods and results of several compelling and recently published articles. We pay special attention to the issue of variable selection. The current literature is divided on the topic of variable selection and we shall include an overview of some literature regarding this topic in this chapter. We contribute to this discussion with the results of our research based on automated feature selection by Adaptive Lasso.

4.1 Credit Losses for Top Down Stress Testing

Stress testing financial institution's profitability or capital condition requires an impact analysis of a stress scenario on the institution's entire portfolio. Most papers that deal with top-down stress testing construct large comprehensive models, based on publicly available data, that predict the capital condition of commercial banks in an adverse scenario. Hence, the focus of these papers is not directly on the prediction of credit losses, but rather on the forecasting of profits and losses on the entire portfolio.

As we mentioned in Section 3.1, it is common to use the NCO rates as a measure for

[1]For a comprehensive overview of stress testing literature we refer to the the first part of the book "Stress Testing the Banking System" by Quagliariello (2009) [21]. For a discussion of top-down stress test approaches for other P&L items we recommend Duane et al. (2013) [35].

credit losses in top-down stress testing. The papers that we discuss below, model the annualized NCO rate process. The models are estimated on regulatory data from the Federal Financial Institutions Examination Council (FFIEC). A great advantage of using NCO data to model credit losses, is that detailed CO, RC, and TL time-series are publicly available for all commercial banks in the United States for several loan categories. In thesis we consider the same data set, which will be discussed further in Chapters 5 and 6.

In their 2012 paper [30] Guerrieri and Welch found that models for P&L items that include macro variables can perform significantly better than a random walk. That is, a model which only takes the current value of an item and an error term into account. In their analysis, they use 22 years of data. They define the following model for each banking measure, or Profit and Loss item $\text{P\&L}_j$ and each macro variable M_i,

$$\text{P\&L}^i_{j,t} = \alpha + \beta \text{P\&L}^i_{j,t-1} + \sum_{k=1}^{4} \gamma_i M^i_{t-k} + \epsilon_t,$$

where ϵ_t is a random variable with mean 0, constant variance, and no auto-correlation. The average forecast of each measure is then constructed by taking the unweighted average of the forecasts of that measure across models,

$$\text{P\&L}_{j,t+h} = \sum_{i=1}^{N} \frac{\widehat{\text{P\&L}}^i_{j,t+h}}{N}.$$

It turns out that this model significantly outperforms a random walk model for net charge offs, and it is concluded that macro variables help explain variation in NCO rates.

Another major contribution is given by Hirtle et al. in their 2014 paper [47]. The results of this paper are enticing, since they offer a thorough set of techniques to make capital projections for institutions in stress scenarios. Hirtle et al. use simple econometric models in a top-down framework to estimate the capital gap of the whole banking industry. The model that is employed to estimate the macroeconomic impact on disaggregated credit losses for individual institutions is chosen by including variables in an auto-regressive model with exogenous variables, based on economical and statistical significance. They use 24 years of data for a panel of 200 banks, and for the NCO rates they disaggregate the loan portfolio and hence the NCOs into 15 subcategories. For each category they estimate,

$$\text{NCO}_{j,t} = \alpha + \beta_1 \text{NCO}_{j,t-1} + \beta_2 M_t + \beta_3 B_{j,t} + \epsilon_{j,t},$$

where M_t is a vector of macro-variables, $B_{j,t}$ is a vector of bank-specific variables for each bank $j = 1, \dots, n_b$, and $\epsilon_{j,t}$ is a normally distributed error term.

Other compelling contributions are by Covas et al.(2014) [43] and Coffinet and Lin (2010) [22]. The former use quantile regression[2] for panel data to account for non-linear dynamics and create asymmetric confidence bands of bank losses. The latter show that macro variables are also relevant for the prediction of French banks' profitability. It turns out that, while most models predict relatively low levels of stress in the banking sector, the quantile approach predicts higher stress levels for the quantiles of the NCO rate distribution.

Our modeling approach is most closely related to that of Kapinos and Mitnik (2015), who use a combination of principle component analysis and Lasso variable selection on a panel of banks to discover the relation between disaggregated P&L items and macro-economic variables that are frequently used in stress testing scenarios. The allure of their method is that Lasso is an automated model selection procedure, allowing for great flexibility in the choice and detail of stress scenarios. Furthermore, the Lasso method shrinks the coefficients of many predictors to zero. One of the most striking properties of this approach is that it allows us to consider more predictors than we have observations. The attractive statistical properties of (Adaptive) Lasso will be discussed thoroughly in Chapter 7.

Kapinos and Mitnik (2015) use a panel of 156 banks for which they remove variation associated with the lag of the response variable by using the following regression,

$$\mathrm{NCO}_{i,t} = \alpha + \beta \mathrm{NCO}_{i,t-1} + \widetilde{\mathrm{NCO}}_{i,t}.$$

They construct a pool of candidate variables K_t based on transformations of macro variables in their stress test scenario. They proceed by selecting a set Z of relevant macro-economic variables from K by using the Lasso method (with dimensions based on the number of observations n and the number of selected variables). Subsequently, they derive the first m principle components of Z and denote these by f_t^m. The effects of these principle components are removed by a regression, and the procedure is repeated for bank-specific variables [54].

4.2 Feature Selection for Top Down Stress Testing

We have presented some existing methods which can be used for the prediction of NCO rates in stress scenarios. The macro-economic and bank-specific variables that were selected by other researchers, may serve as a guideline for the features that we consider. A disadvantage of the current literature is that the availability of results concerning the evaluation of methods in terms of in-sample and especially in terms of out-of-sample performance is limited.

[2] Quantile regression is a technique to estimate all quantiles of a probability distribution. See [1] for more information.

Guerrieri and Welch (2012) model aggregated NCO rates and find that gross domestic product, unemployment rate and the house price index give the best performance vis-à-vis prediction accuracy [30]. Hirtle et al. (2014) disaggregate the net charge offs in 15 categories, such as "First Lien Residential Real Estate Loans" or "Credit Card Loans". The explanatory variables that were selected differ per category and are given by change in unemployment, home price growth, home price growth if it is negative (zero otherwise), commercial property price growth, commercial property price growth if it is negative (zero otherwise), and a time trend [47]. Variables were selected based on economic significance and measures of in-sample performance. Other considerations include consistency with economic theory and previous empirical work. An example is that theory and research suggest that loss rates on mortgages rise in a convex way as property prices decline. Therefore they include the nonlinear transformations of the commercial property and house price index.

Kapinos and Mitnik (2015) use automated selection and principle component analysis to select the ultimate specification of their model for NCO rates for all loans and leases. The macro-economic variables that they consider are squares, cubes, and lags of variables that are specified in the stress tests designed by the Federal Reserve[3]. It turns out that the Lasso method does not select financial indicators such as yields and interest rates for the NCO model. They also select bank specific variables such as the percentage of consumer loans, the amount of past due loans, and the value of held-to-maturity securities [54].

A contribution of this thesis is that we include a discussion on the influence of aggregating or disaggregating the loan portfolio for the estimation and forecasting of net charge off rates. This gives a new dimension to the modeling of credit loss rates. Our results also show that there are (transformations of) macro-economic variables that can be relevant for the prediction of NCO rates that are not included in the models used in current literature. We provide an evaluation of the strengths and weaknesses of the chosen method and compare several modeling options based on their predictive strength. We could not find publications with a thorough evaluation of the performance of top-down stress testing models for credit losses. Furthermore, we shed light on the limitations of (top-down) stress testing and propose new concepts that can alleviate these limitations. This can open the path for a top-down stress method of this type that is applicable to Rabobank.

4.3 Automated Feature Selection

There exist many strategies for the construction and subsequent selection of models. Depending on the application, some of these strategies might be more suitable than

[3]The Federal Reserve stress test scenario will be discussed in more detail in Chapter 5.

others. The issue that we want to raise here is the matter of the advantages and disadvantages of automated feature selection techniques as in [54], compared to a more classical approach of selecting a model based on expert opinion or theoretical relations as in [47]. For the bottom-up approach one considers effects on a micro-level and theoretical models are imperative. But the exact specification and workings of the macro-economy are not well understood, making it challenging to construct top-down models for credit losses. We pay special attention to literature on feature selection for macro-economic top-down applications.

In Section 4.2 we have seen that in most of the existing papers on top down stress testing features are selected based on expert opinion and a variety of statistical in-sample criteria. Kapinos and Mitnik (2015), on the other hand, use an automated feature selection procedure [54] and use the Lasso method. We follow their approach, in the sense that we use Adaptive Lasso as an automated supervised learning method to find a suitable model for the prediction of NCO rates conditional on macro-economic developments. In this thesis we motivate the advantages of using the adaptive Lasso to automatically select relevant features in design matrices based on typical macro-economic stress scenarios. We take the approach of Kapinos and Mitnik [54] one step further and consider even more predictors than they did. We use Adaptive Lasso to simultaneously select and estimate a parsimonious model from a large set of data. A major advantage of such an approach is that our method can easily be adapted to be suitable for different stress scenarios.

Much ink has been spent on the topic of automated selection procedures. We summarize the discussion in the Elements of Statistical Learning [19]. In a classical research setting, a researcher has conjectured an hypothesis, designs an experiment, and uses statistical methods to test the hypothesis. In the case of automated feature selection, the data itself is used to discover a hypothesis. Using such a discovered hypothesis directly, may lead to overly optimistic results, due to over-fitting and selection bias. To circumvent bias and over-fitting, the dataset must ideally be split in a training, validation, and test set. The training set is then used to learn the features and estimate the model. The validation set is used to determine the prediction error for model selection. And finally, the test set is used to determine the generalization error of the chosen model.

An example of statistical learning, as described above, in macro-economic modeling is given in an article by Epprecht et al. (2013), who show that Lasso methods are superior to classical models in forecasting Gross Domestic Product (GDP) in the United States [36]. Doornik & Hendry (2015) describe how our statistical reach, in relation to macro-economics specifically, can be extended to the discovery of new knowledge by machine learning methods with automated feature selection procedures such as Lasso [51]. In Chapter 8 we detail how we apply their approach to the top-down stress testing estimation problem that is central to this thesis.

An interesting paper that uses Lasso to identify determinants of retirement in Denmark is by Kallestrub-Lamb et al. (2013) [38]. Kock and Teräsvirta (2014) show how automated selection methods can be used to forecast macro-economic variables in the 2007-2009 crisis [48]. For an comprehensive overview of sparse high-dimensional models in economics we refer to Fan et al. [27]. The application of Lasso methods to macro-economic data is relatively scarce in literature. Another contribution of this research project to existing literature is the establishment of a new application for Adaptive Lasso methods.

Part II

Data

Chapter 5

Available Data

Our main research goal is to assess the effectiveness of the prediction of NCO rates conditional on exogenous paths of macro-economic variables. To that end, we wish to disaggregate the loan portfolio of an individual bank or the averaged loan portfolio of all banks, to a suitable level of granularity to capture the impact of specific stress scenarios. Subsequently, we require a model for NCO rates conditional on macro-economic indicators for each loan category in the chosen aggregation level. It can be constructive to include bank-specific variables that characterize an individual bank to explain heterogeneity of NCO rates between banks.

Based on the literature study on top-down stress testing, we speculate that it is prudent to use sufficiently long time-series on disaggregated loan loss data. Contrary to the availability of data for European banks, there exists a large publicly accessible data source for banks in the United States. In this chapter we specify which data is available for a stress testing exercise across commercial banks in the United States. We consider disaggregated net charge off data, macro-economic data, and other bank-specific data.

5.1 Net Charge Offs

The primary data-sources for banking variables in the United States, including NCO rates, are the publicly available *call reports*[1]. All banks with offices in the United States are obligated to file a detailed call report of their income statement and balance sheet to the Federal Financial Institutions Examination Council (FFIEC). These call reports contain accounting fields, specified by the FFIEC or Federal Reserve. A major advantage of this dataset is that the reported values of different

[1] Call report data from 1976 to 2010 can be retrieved from `https://www.chicagofed.org/banking/financial-institution-reports/commercial-bank-data`. Call report data starting from 2011Q1 can be found at `https://cdr.ffiec.gov/public/`

banks are comparable.

The call reports contain, among other items, disaggregated reported values for Charge Offs (CO), Recoveries (RC), and Total Loans (TL) for all banks[2]. By equation 3.2 these can be used to calculate the NCO rates for several loan categories. This data is available for a large number of banks; the amount of unique commercial bank identification numbers in the call reports between 1990 Q2 and 2015 Q2 is 17803.

The charge offs and recoveries reported in the first quarter refer to those that are incurred during the first three months of the year. In Table 5.1 below we see the time periods for each quarter of the year. The TL value is quoted twice in the call reports, based on the quarterly average or the end-of-period value.

Table 5.1: Quarterly Data

Q1	January 1$^{\text{st}}$ - March 31$^{\text{st}}$
Q2	April 1$^{\text{st}}$ - June 30$^{\text{th}}$
Q3	July 1$^{\text{st}}$ - September 30$^{\text{th}}$
Q4	October 1$^{\text{st}}$ - December 31$^{\text{st}}$

We use equation 3.2 to construct time-series of $\text{NCO}_{i,j,t}$ rates for each loan category $P_i \in \mathcal{P}$ and each bank $j \in \{1, \dots, n_B\}$. We consider $t = 1, \dots, n_T$, where the first time-point refers to 1991 Q1 and n_T denotes 2014 Q4. In Figure 5.1 the structure of loan categories as reported on the call report of the 2014 Q4 forms is displayed. In Chapter 6 we discuss the dataset and the changes in this loan structure over time.

[2]Aggregated net charge off data can be found at `http://www.federalreserve.gov/releases/chargeoff/chgallnsa.htm`

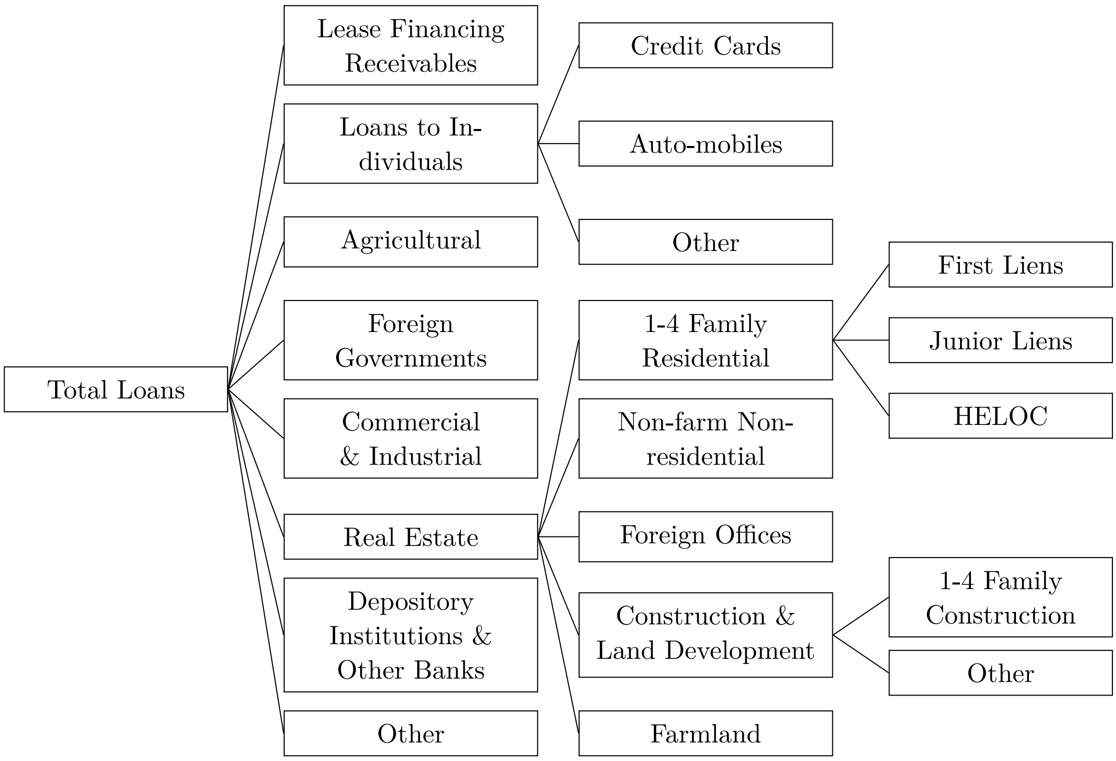


Figure 5.1: Aggregation of the loan portfolio in Federal Reserve call reports for 2014 Q4. Information was extracted by analyzing the publicly available FFIEC data.

5.2 Macro-Economic Variables

Large banking institutions in the United States are required to perform annual stress tests under the Dodd-Frank Act and submit the results to the supervising authority; the Federal Reserve. Consistent with the approach in the literature, we consider all macro-economic variables that are provided by the Federal Reserve in their Dodd-Frank Act Stress Testing (DFAST) stress test scenarios as a starting point.

The Federal Reserve uses 16 domestic variables and 12 international macro-economic indicators to describe baseline, adverse, and severely adverse stress scenarios for financial institutions. The domestic indicators that they consider are quarterly time-series of the GDP, inflation, household's disposable income, yields, interest rates, stock prices, house prices, commercial property prices, and the volatility on the markets. International variables include GDP growth, inflation and exchange rates for the Euro area, developing Asia, Japan, and the United Kingdom. Shortly, we will discuss the meaning of these stress indicators. The stress horizon that the Federal

Reserve sets for its stress testd is 9 quarters, where typical stress horizons are usually between 8 and 20 quarters.

The Federal Reserve stress test scenarios are designed to stress capital ratios in specified adverse scenarios. For such an exercise, one not only stresses credit losses, but other P&L items as well. For the prediction of net charge off rates conditional on a macro-economic scenario, we do not consider financial indicators such as yields and interest rates here. Instead, we focus on the Federal Reserve indicators that are associated with real economic growth and price indices. This approach is in agreement with the top-down stress testing literature, as was discussed in Chapter 4.

We give a brief description of these macro-economic variables and the source of the data below.

1. The level of house prices in the United States can be measured by the House Price Index (HPI). We use the seasonally adjusted Standard & Poor/Case-Shiller U.S. National Home Price Index, with ticker CSUSHPISA from the FRED database[3]. This index tracks the value of single-family house prices in the United States.
2. The value of commercial property can be represented by the Commercial Property Price Index (CPPI), which is published by the National Council of Real Estate Investment Fiduciaries (NCREIF) and can be found on the website of NCREIF. We calculate the index from the reported quarterly returns on commercial property by increasing the level with percentages equal to the return for each quarter.
3. The Dow Jones Industrial Average (DJIA) represents overall market performance. It is calculated as a scaled and weighted average of stock values of 30 large publicly traded companies that are perceived as being representative of the industrial sector. We use data that is available at the online Yahoo financial database under the ticker $\hat{\text{D}}$JI.
4. The Chicago Board Options Exchange Market Volatility Index (VIX) is a popular index that tracks implied volatility for S&P 500 options. In this thesis we use the CBOE volatility index, available online in the FRED database, it can be found with the ticker VIXCLS.
5. In order to represent inflation we use the Consumer Price Index (CPI). The CPI is a measure of the change in the prices of goods and services paid by urban consumers. We use the seasonally adjusted consumer price index for all urban consumers, with ticker CPIAUCSL, from the FRED database.

[3] Many financial and economic time series can be retrieved from the publicly available FRED database: `https://research.stlouisfed.org/`.

6. Disposable Income (DI) is frequently used as a measure to gauge overall economic performance. It represents the amount of available money that households have for saving and spending after taxes have been taken into account. As a measure we take seasonally adjusted real disposable income rates per capita, from the FRED database, which can be found with ticker A229RX0Q048SBEA.
7. The Gross Domestic Product (GDP) is the leading macro-economic indicator and is often used to gauge the performance of a country's economy. It represents the total amount of goods and services produced in a certain period of time and as such, it can be interpreted as a measure of the size of the economy. The data we use are the seasonally adjusted rates for real gross domestic product in the FRED database, under the ticker GDPC1.
8. The Unemployment Rate (UR) represents the portion of the population that is currently unemployed, but actively seeking employment. The unemployment rate is generally seen as a lagging indicator that confirms longterm market trends. We use the seasonally adjusted unemployment rate, which is also available in the FRED database and can be found with ticker UNRATE.

In Figure 5.2 the time-series for all these indicators between 1991 Q1 and 2014 Q4 are displayed. We apply transformations to the macro-economic time-series so that the data is approximately stationary. We say that a process x_t is weakly stationary when,

i $\mathbb{E}[x_t] = \mu$, $\forall t$,

ii $\mathrm{Var}(x_t) = \sigma_x^2$, $\forall t$,

iii $\mathrm{cov}(x_t, x_{t-s}) = \sigma_s$, $\forall t, s$.

Most econometric modeling techniques require stationary predictor and response variables. Stationarity in econometric time-series modeling is discussed further in Appendix A. The time-series for CPPI, GDP, DJIA, DI, CPI, and HPI are adjusted by taking the logarithm of the values and then differencing the series. This is a common and frequently used method to transform non-stationary data. Since UR and VIX are already roughly stationary, these variables are not transformed. From here on, we shall only refer to the (approximately) stationary time-series of the macro-economic variables.

Often, reported values for macro-economic indicators are adjusted at a future date, when more accurate data has become available. The data as it was known at an earlier point of time is called *vintage* data. In forecasting exercises, this vintage data usually gives more reliable estimates of the model's prediction accuracy. However, in a stress test we assume that the future as specified by the stress scenario is known exactly, and therefore we choose not to use vintage data. The data in Figure 5.2 is

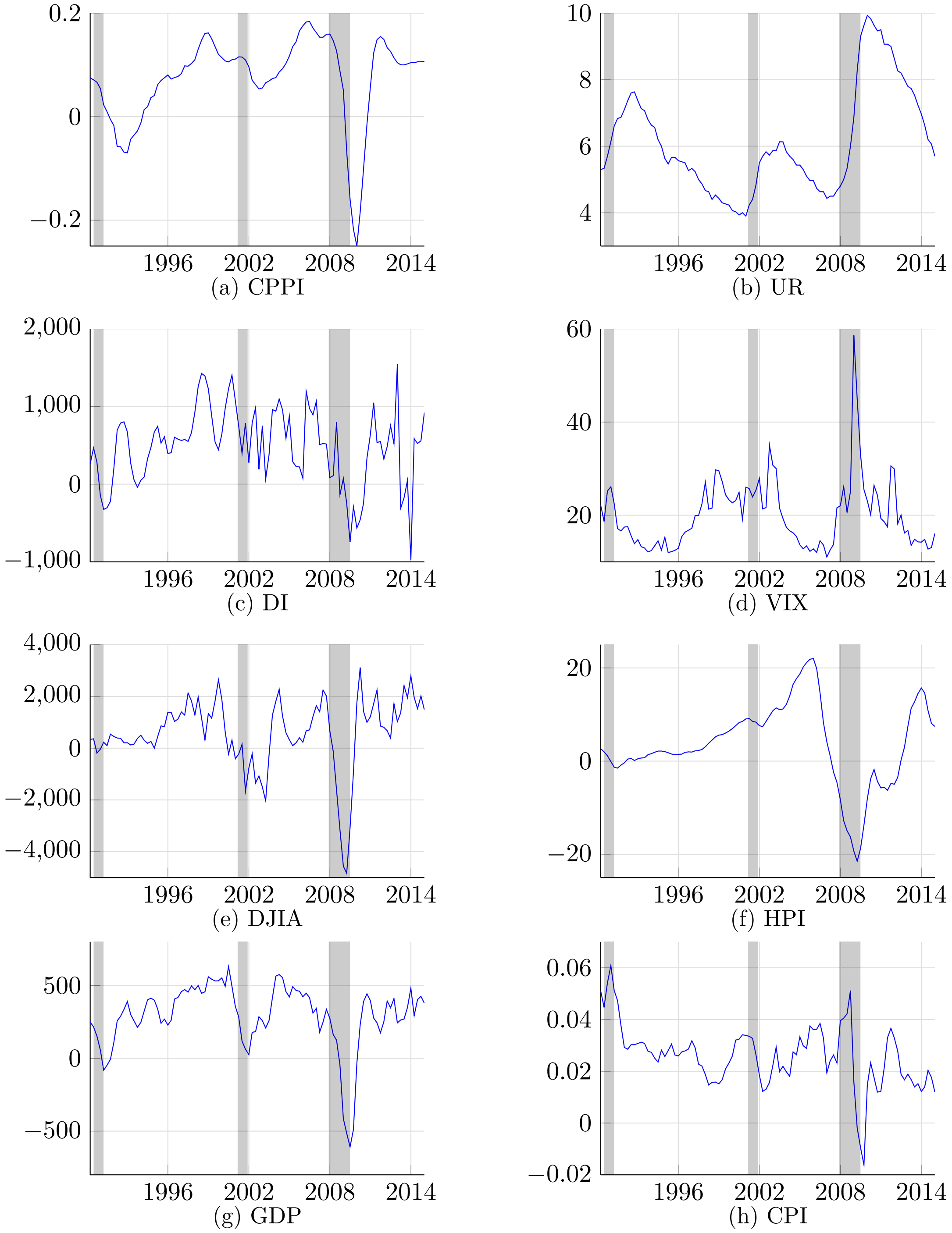


Figure 5.2: Non-financial, domestic macro-economic variables that are used in the Federal Reserve's stress test scenarios. Data was collected from the FRED database, Yahoo Financial Database and the NCREIF website. The shaded areas represent the recession periods as determined by the National Bureau of Economic Research (NBER).

of the 2015 Q1 vintage point.

In our final model we want to make use of *indicator* functions as in Hirtle et al. [47]. Such a function for a macro variable M_t^1 can be denote as $\mathbb{1}_{\{M_t^1 > m\}}$ and is equal to one if $M_t^1 > m$ and zero otherwise. Such functions can capture convex behavior. For Unemployment Rate in the United States, the Federal Reserve publishes the Natural Rate of Unemployment (NROU), which can be found under the ticker NROU on the Federal Reserve website. The natural rate of unemployment is supposed to measure the expected Unemployment Rate *excluding* cyclical factors. Or rather, the rate of unemployment as if there were no boom and bust periods. We use this rate to construct the indicator function $\mathbb{1}_{\{\mathrm{UR}_t > \mathrm{NROU}_t\}}$, which indicates whether the unemployment rate is higher than expected at time t.

Indicators for the other macro-economic variables that we consider are constructed by evaluating whether its value is larger than zero. For the VIX, we compare the series to its long-term mean, where we can also use the future data in the stress scenario for the averaging since it is assumed that the macro-economic future is known, and only the bank losses need to be estimated.

5.3 Bank-Specific Variables

One of our research goals is to learn the role of bank heterogeneity and banking variables in the estimation and prediction of NCO rates. The literature study showed that the composition of the loan portfolio can play a crucial part in determining an individual bank's credit loss rate. But there might be other relevant bank-specific factors that influence the NCO rate. In this section, we discuss which variables we include in our model to estimate possible structural individual effects.

Therefore, we first want to determine potentially relevant bank-specific characteristics or variables for the prediction of banks' NCO rates. Net charge offs do not only vary over time as a function of the macro-economic variables in the stress scenario, but they may also vary across banks. A panel data model allows us to capture heterogeneity between banks, or to differentiate between the impact of an economic shock on credit losses based on banking variables.

Broadly speaking, the impact of a macro-economic stress scenario on an individual bank depends on how a bank manages its assets, liabilities and equity. Other variables such as geographical location, risk appetite, capitalization or variables associated with idiosyncratic effects, may also explain some of the variance in NCO rates for individual banks. Depending on the application of the stress test, it can be more informative to only take system wide events into account. However, the severity of the impact of a systemic shock can depend on the quality of the loan portfolio or management thereof.

As bank-specific explanatory variables B_t, we consider the variables that are given by the CAMEL criteria, which are displayed in Table 5.2. These are widely accepted measures for the assessment of the quality of a bank [31]. Additionally, CAMEL is a frequently used international bank rating system employed by supervisory institutions and researchers. The CAMEL measure includes statistics on the capitalization, asset quality, management quality, profitability and liquidity. Hirtle and Lopez (1999) [4] found that past CAMEL ratings gave useful insight into current bank conditions. The effect of past ratings was significant up to 12 quarters.

In Table 5.2 financial measures that can be indicative of the CAMELS criteria are given in the second column. The third columns gives a definition of these financial measures in terms of accounting fields that can be extracted from the call reports that we introduced in Section 5.1. Detailed information about the derivations of these time-series from the call reports is provided in Table D.2 in Appendix D.

Table 5.2: CAMEL criteria

Description	Financial Measure[1]	Definition
Capital Adequacy	Capital Ratio	Total Capital / Total Assets
Asset Quality	Provision Rate	Loan Loss Provision/ Total Loans
Management Quality	Non-Interest Profit	Non-Interest Expense / Operating Income
Earnings	Return on Equity	Net Income / Total Capital
	Net Interest Margin	Net Interest Income / Total Assets
Liquidity	Total Liquid Assets	Sum of liquid assets
Size	Total Loans, Loan growth	Total Loans, Total Loans (% change)

[1] The CAMEL(S) criteria are expressed in terms of financial ratios. The ratios that are presented here, were derived from V. Behbood (2012) [29]. We also consider loan size and loan growth as financial measures for 'size'.

We extend the CAMEL criteria with measures that represent the size and growth of the loan portfolio. We refer to this extended version as CAMELS criteria. These variables potentially help explain the variance among NCO rates of individual banks. We discard asset quality from the set of measures, since NCO rates themselves can be used as a measure for the quality of assets.

Chapter 6

Data Handling

The Federal Reserve dataset, consisting of quarterly call reports for commercial banks poses challenges when it is used for modeling, because the framework was primarily designed for regulatory purposes and not for data analysis. A second aspect to take into account is that effects caused by decision-making within banks, are present in the data. This makes a sound handling of the data an important task. In this chapter we describe how the data was collected, selected, and adjusted to optimize the dataset. Furthermore, we consider the problem of disaggregating the loan portfolio into categories $P_i \in \mathcal{P}$ of loan portfolios, based on data-availability.

6.1 Forming Consistent Time-series

Defining variables, which are derived from accounting fields on the call reports, correctly and consistently over time is a daunting task, because definitions and reporting rules have occasionally changed considerably. For the construction of a consistently defined time-series we also need to consider the handling of irregularities due to mergers between banks and other problematic observations, such as negative Net Charge Offs rates. Below we discuss our adjustments to the data and the selection criteria we used to circumvent irregularities in the data. Our approach closely follows that of Kashyap and Stein, and De Haan et al. [6, 56].

6.1.1 Extracting Loan Categories and Aggregation Levels

We downloaded the call reports for all banks with domestic offices in the United States for each quarter from 1990 Q2 to 2015 Q2. Each call report consists of a few thousand accounting fields, which can be found on the balance sheet, P&L statement, or other financial statements[1]. The Federal Reserve Bank of Chicago

[1] The call report form for 2014 Q4 can be found at: `http://www.ffiec.gov/pdf/FFIEC_forms/FFIEC031_201412_f.pdf`

also provides a data dictionary on its website[2], which contains definitions, changes in these definitions, and general information about the reporting structure of each item. As a general rule, we only consider variables on the call reports which are similarly defined between 1990 Q2 and 2015 Q2.

According to the Federal Reserve's data dictionary, some of the time-series for loan categories are not available for the entire period 1990 Q2 to 2015 Q2. For other categories their definitions have changed during this period. Consistently defined time series can be constructed for a maximum of 12 loan categories in the following way. First, we consider the lowest level of aggregation of the loan structure of 2015 Q2, which is identical to the hierarchy displayed in Figure 5.1. If a category was not specified for the entire period, or its definition has changed during this time, we move up one level. We iterate this process until all categories are consistently defined. The resulting ordering of loan portfolios is displayed in Figure 6.1.

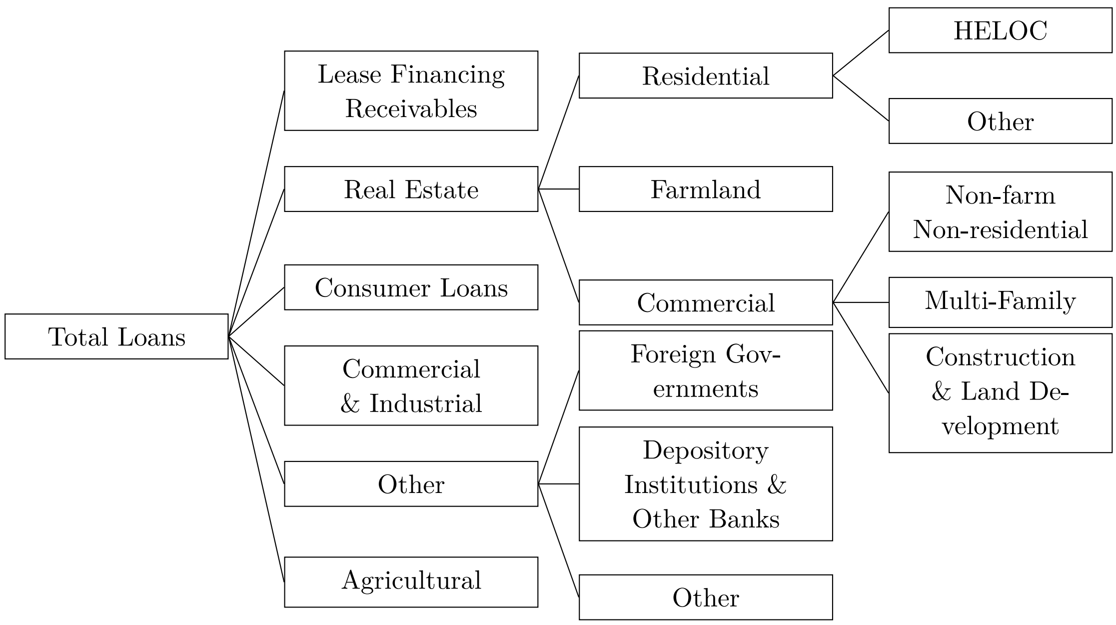


Figure 6.1: Consistent Loan Categories in Call Reports 1990 Q2 to 2015 Q2.

Depending on the type of bank either a FFIEC/031 or FFIEC/041[3] is filed quarterly. The former call report is filled out by banks with both domestic and foreign offices whereas the latter is filled out by banks with domestic offices only. In the FFIEC/031 reporting form a distinction is made between time-series on a consolidated basis[4]

[2] `http://www.federalreserve.gov/apps/mdrm/data-dictionary`

[3] Until 2000 Q4: FFIEC/032, FFIEC/033 and FFIEC/034 reports.

[4] A consolidated basis means that the reported values in the accounting fields are given for the

and time-series for accounting fields on a domestic basis. This is the case for some fields regarding loan and deposit data. They are referred to as the RCFD and the RCON series, respectively.

Generally, banks with foreign offices provide data on a consolidated basis only. For banks with no foreign offices, it can be assumed that the RCON en RCFD series are identical [56]. Because we are interested in determining the effect of the developments in the macro-economy of the United States, it is sensible to consider domestic loan variables. We follow De Haan et al. [56] and use the RCON series for loan and deposit data. For the bank-specific variables in Section 5 domestic values are not published by commercial banks with foreign offices. Therefore we use the RCFD series for all other variables.

A detailed description of the selected loan categories $P_i \in \mathcal{P}$ and the code, with which their respective time-series can be extracted from the call reports, is provided in Table D.3 in Appendix D.

We use the data in Figure 6.1 to construct aggregation levels A_1, A_2, and A_3 for the estimation and prediction of NCO rates. The levels that we consider are constructed considering the categorization in the call reports, economical and statistical similarity between categories, and the typical relative size of the loan category compared to the total loan portfolio. NCO rates for combined loan categories can be easily obtained by recalculation with equation 3.9. Intuitively, higher aggregation levels may reduce the amount of noise that is due to idiosyncratic events or other unexplained effects. On the other hand, it may obscure the effect that macro-economic developments have on specific parts of the loan portfolio. Also it reduces the specificity in rates for individual banks that we obtain by taking differences in loan composition into account.

The aggregation levels are displayed in Table 6.2. We will use these levels to discover and estimate models that predict NCO rates conditional on macro-economic variables. The forecasts can be combined to obtain predictions for the credit loss rates of individual banks. This procedure will be introduced and explained in Section 8.3.

The highest level of aggregation, or A_3 consists solely of the loan category Total Loans. The middle aggregation level A_2 is made up of Consumer (CON), Commercial Real Estate (CRE), Residential Real Estate (RRE), Commercial & Industrial (C&I) & Leases on Financial Receivables (LEASE), and Other (OTHER) loans. The lowest level of aggregation consists of the loan categories Consumer (CON), Non-farm Non-Residential (NFNR), Multi-Family (MF), Construction & Land Development (CLD), Home Equity Lines of Credit (HELOC), Closed-End Residential Real Estate (RES), Agriculture (AGRI), Farmland (FARM), Commercial & Indus-

combination of the domestic and foreign portfolios.

trial (C&I), Leases on Financial Receivables (LEASE), and Other (OTHER) loans. Note that the definition of OTHER loans is different for aggregation levels A_2 and A_1.

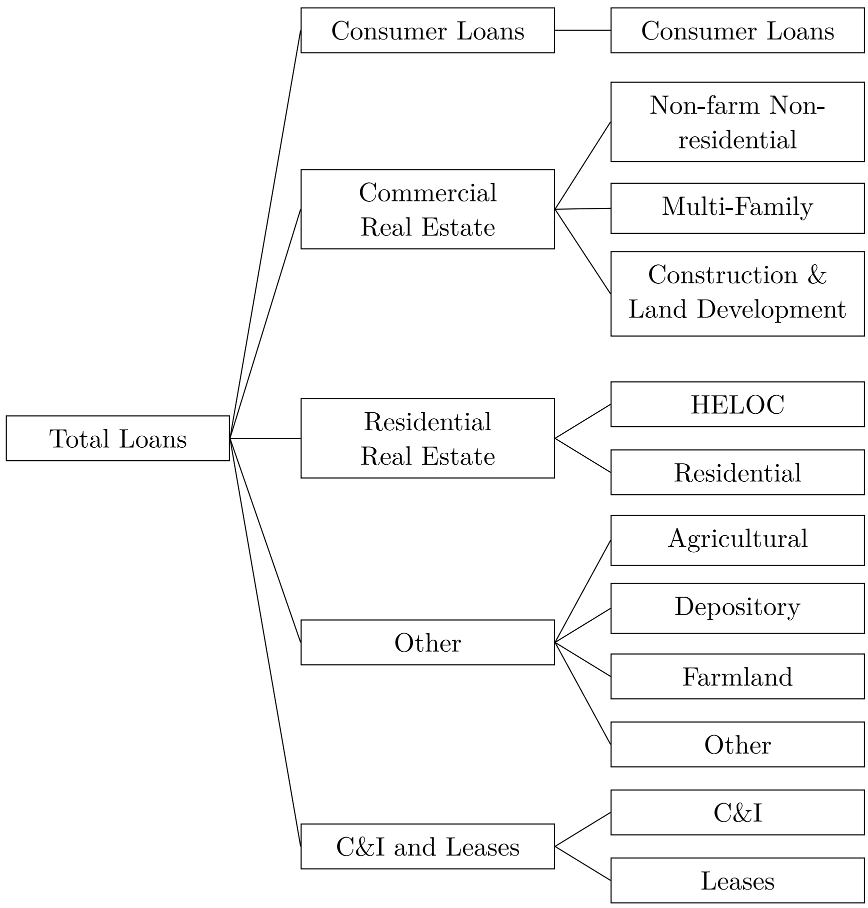


Figure 6.2: The aggregation levels A_3, A_2, and A_1 are displayed from left to right. Note that Farmland Loans are actually categorized under real estate in the call reports, but in level A_2 we consider it as a part of the category Other.

6.1.2 Accounting for Mergers

Mergers occur frequently in the banking industry. Since a merger involves the transfer of a large part of the loan portfolio of one id-number to another, mergers may lead to a significant distortion of the balance sheet and income statement data that is reported to the FFIEC in the quarter of the merger and the subsequent period [6, 56].

Merger data was extracted from the website of the Federal Reserve Bank of Chicago[5]. In order to avoid irregularities or inconsistencies in the time-series on the individual bank level due to mergers, we follow [56] by applying the following *screens* to our data. The fields non-survivor id, survivor id, and merger dates in the merger data file, were used to perform the first screen listed below.

1. Banks involved in a merger are dropped from the observations in the quarter of the merger and the following quarter.
2. Observation t of bank j is dropped if the loan growth is more than five standard deviations away from the cross-sectional mean growth.
3. Observation t of bank j is only included when the previous two loan growth variables are available.

The first screen drops observations for merging banks. It is also possible that a large transfer of assets takes place, but both reporting banks continue to exist. The second screen filters out those observations for which a large transfer of assets has taken place. The last screen filters out observations of loan portfolios that have recently undergone a substantial change in size.

6.1.3 A Closer Look at NCO Time-Series

Commercial banks file their call reports under a FFIEC id-number, which we can use to obtain its financial report for each quarter from the total set of reports. This is a convenient way to construct individual banks' time-series. However, the distinction between commercial banks and the identification numbers has some issues that we must keep in mind.

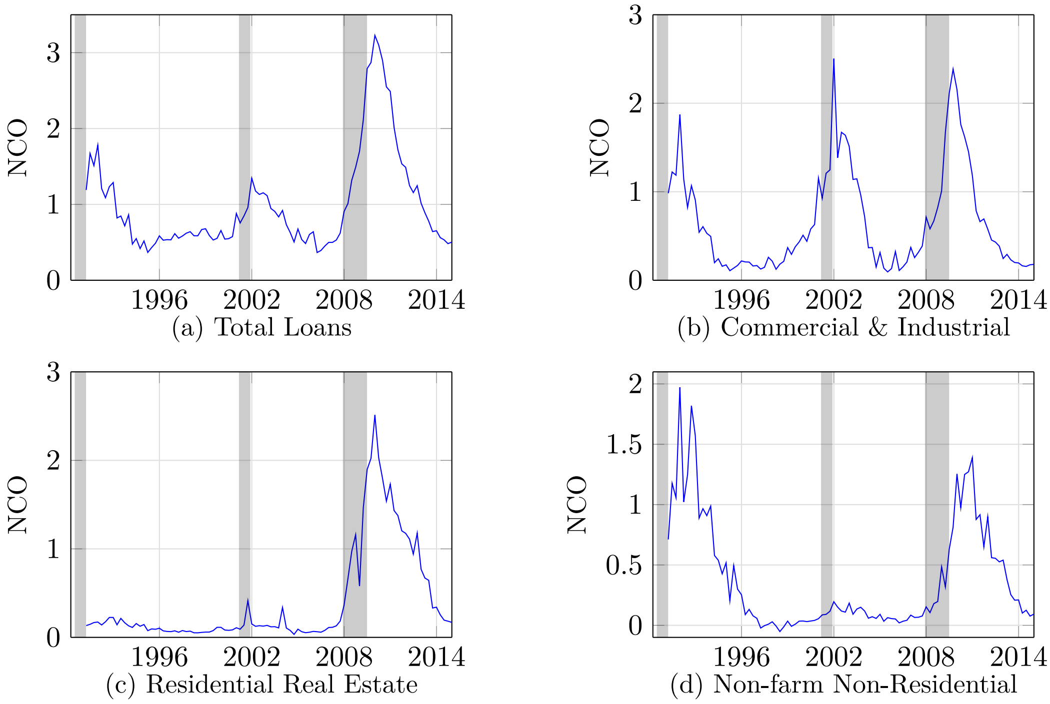
Some banks, for example Rabobank North America, have filed call reports under several id-numbers. Thus, the reported values of the accounting fields of these banks typically do not correspond to the values that can be found in quarterly reports. Also, especially for id-numbers that report a relatively small amount of assets, it occurs that there are substantial changes between the reported data in consecutive quarters. Keep in mind that this may be a source of irregularities that can have an adverse effect on any estimation procedure. Another aspect is that the call reports are filed by commercial banks only, and therefore investment banks are not included. Hence, the available time-series for some large banks such as Morgan Stanley, which was an investment bank before 2007, are relatively short and were not included previously.

Besides the effect of these irregularities, the differences in mean, variance and shape

[5] Files can be downloaded from: `https://www.chicagofed.org/banking/financial-institution-reports/merger-data`

of the NCO rate series can differ greatly between banks. In order to capture the systemic macro-economic relation, we consider NCO data of multiple banks at once. For our first method, we therefore use averaged time-series for the NCO rate. Our second method uses a panel data model, to separate the bank-individual effects from the systemic effects.

Recall that average NCO rates refer to the NCO rate of all the loan portfolios of reporting banks combined. We construct averaged time-series by loan category, by adding the CO, RC, and TL for all commercial banks that report to the FFIEC, and subsequently, we calculate the rate. The resulting time-series for the NCO rate is displayed in Figure 6.3, for some loan categories that can be found in the hierarchy in Figure 6.1.



(a) Total Loans (b) Commercial & Industrial (c) Residential Real Estate (d) Non-farm Non-Residential

Figure 6.3: Average Net Charge Offs rates for the loan categories Total Loans, C&I, RRE, and NFNR as reported on the call reports. The shaded areas represent the recession periods as determined by the NBER.

Note that the loan categories Total Loans and Commercial & Industrial (C&I) Loans display three clear peaks in the period 1991 Q1 to 2014 Q4. These seem to coincide roughly follow the recession periods. The categories Residential Real Estate (RRE) and Non-farm Non-Residential (NFNR) were stressed one and two times, respectively. The main idea of stress testing a loan portfolio on a disaggregated level is to link the macro-economic variables displayed in Figure 5.2 to disaggregated NCO rates, displayed in Figure 6.3.

6.1.4 Other Irregularities in the Data

The NCO rates that are derived from the call reports using equation 3.2 are approximations to the actual rate at that time. The main issue with this is that recoveries on loans that were charged off, appear on the call reports in other quarters then the charge off itself. This may result in negative NCO rates for some quarters. Negative rates are encountered even when the charge offs, recoveries, and loans are averaged over all commercial banks in the data set. Since other costs are involved when loans are written off, NCO rates above 100% can also occur.

Another issue is that a bank's loan portfolio is finite and therefore there can be quite some variability on the charge offs, recoveries or loan total over a relatively short time period. For smaller loan portfolios it can happen that there are no credit losses at all in some quarters, which is of course not representative of the true value of the NCO rate. This issue is alleviated by only considering the 100 banks with the largest loan portfolios. Including smaller loan portfolios in the estimation procedure, requires methods that can handle many zero observations.

Other irregularities that we encountered were that some values were not reported or that negative values were reported in accounting fields that should be positive, such as capital. In a few cases, some of the financial ratios of the CAMEL(S) criteria in Table 5.2 rendered extremely unlikely values. When the values for financial ratios were extraordinarily high, we classified the observation as non-representative and excluded it from the panel. Below we list the additional screens that we applied to the data. Observation t of bank j is excluded when:

- i The NCO rate exceeds 100%.
- ii The capital ratio exceeds 50%.
- iii The return on equity exceeds 100% or is lower than -100%.
- iv One of the CAMEL(S) criteria or the NCO rate is not reported.
- v The value for accounting field that should be positive, is negative.

The first screen is applied because NCO rates higher than 100% are theoretically speaking not sensible, since the bank cannot lose more than it has loaned out. The second and third screen are applied to avoid including observations of special banks that report extreme capital ratios or return on equity. If our model is not able to fully explain such variations, this can have a large impact on our estimation. Understanding this type of behavior is not integral to the prediction of credit loss rates for large banks. The amount of observations that are removed due to these two screens for the 100 largest observations over the quarters of data that were included in our training set was 6. The last screen is applied to ensure that all observations in the panel are complete, which is necessary for the selection and estimation procedure.

6.2 Filtering Time-series

After the application of the screens proposed in Sections 6.1.2 and 6.1.4 the dataset is not directly suitable for econometric modeling purposes. The purpose of this section is to explain other adjustments that we made to the data.

The charge offs and recoveries on loans and leases are reported as year-to-date on the call reports. Following Hirtle et al. [47] and Kapinos et al. [54], we consider annualized net charge off rates. These can be retrieved from the data by taking the difference of charge offs and recoveries with their value in the previous quarter, except for call reports in the first quarter. Annualized rates are then obtained by multiplying equation 3.2 with $4 \cdot 100\%$. This can be written down as follows,

$$\begin{aligned} \mathrm{NCO}_t &= 4 \cdot \frac{(\mathrm{CO}_t - \mathrm{CO}_{t-1}) - (\mathrm{RC}_t - \mathrm{RC}_{t-1})}{\mathrm{TL}_t} \cdot 100\% && \text{if } t \text{ is in quarter } 2, 3, \text{or } 4, \\ \mathrm{NCO}_t &= 4 \cdot \frac{\mathrm{CO}_t - \mathrm{RC}_t}{\mathrm{TL}_t} \cdot 100\% && \text{if } t \text{ is in quarter } 1. \end{aligned}$$

The NCO rate data that can be obtained in this fashion contains substantial effects caused by decision-making in reporting, especially for individual banks. We observe that the rates can vary a great deal between quarters. Possibly this happens because there is no guidance as to when the charge offs (CO) an recoveries (RC) should appear on the income statement. Harris et al. (2015) [53] note that seasonality may affect quarterly data, concerning credit losses, both for accounting and economic reasons. An example is given by [3], in which it is shown that loan loss provisions are often delayed to the last fiscal quarter, when the audit usually occurs. Harris et al. propose to average the data by,

$$x'_t = \frac{1}{4}\left(x_t + x_{t-1} + x_{t-2} + x_{t-3}\right), \tag{6.1}$$

to circumvent issues arising from the reporting error on the actual loss rate.

However, since x'_t in equation 6.1 is the average of the NCO rates in the previous 4 quarters, the NCO rate becomes a lagging variable. We propose to use a centered moving average instead and define for a time series of length n_T

$$x'_t = \frac{1}{8}x_{t-2} + \frac{1}{4}x_{t-1} + \frac{1}{4}x_t + \frac{1}{4}x_{t+1} + \frac{1}{8}x_{t+2}, \qquad \text{for } t = 3, \ldots, n_T - 2, \tag{6.2}$$

and exclude the first and last two observations of the time-series. Note that for a viable application to Rabobank, we would like to use a more advanced filtering method to remove the seasonality and reporting errors during the fiscal year. Firstly, because discarding four data points is costly, and secondly because this method is quite crude and could eliminate important information in the data.

In econometric modeling it is often assumed that time-series can be decomposed into four components: trend, cycle, seasonality, and white noise. This is commonly referred to as classical decomposition, and the general idea is presented in appendix A. The seasonal averaging in equation 6.2 is a simple method to estimate the trend-cycle in the NCO rate data. Ideally, it removes seasonality, seasonal reporting errors, and other noise factors from the data. For our top-down stress test model, we are mostly interested in the variation over a time span of 8-20 quarters, and since we are interested in long-term effects we analyze the trend-cycle component instead of the reported data. The result of a five-point moving average transformation is illustrated on averaged NCO rates on *farmland* loans in Figure 6.4.

Even after applying the screens that were described in Sections 6.1.2 and 6.1.4, some of the NCO rates were negative, both on the aggregate and individual level. This is likely due to the discrepancy between the reporting dates for charge offs and recoveries on a loan which has gone into default. An advantage of the seasonal-averaging filter is that it reduces the amount of negative rates greatly. An example of this can be observed in Figure 6.4.

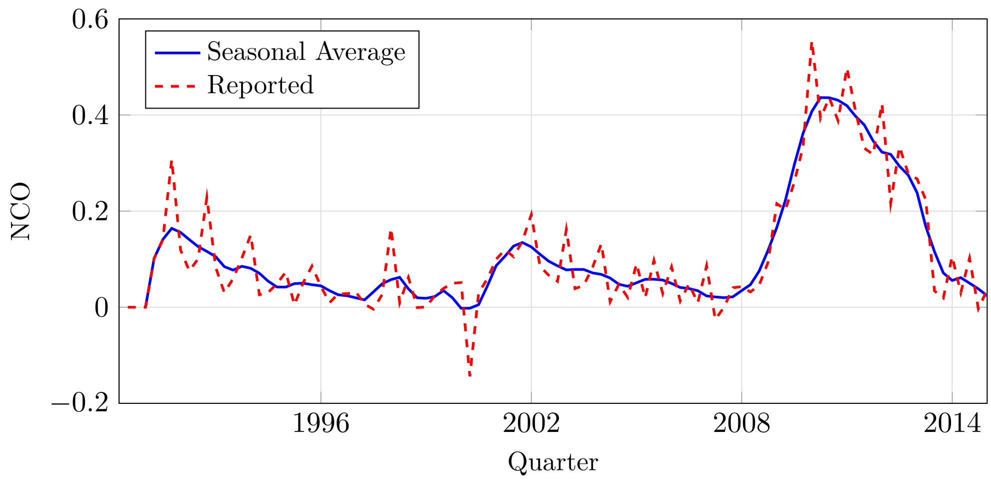


Figure 6.4: Seasonally averaged and reported net charge off rates for the loan category "Loans secured by Farmland" on the average banking level.

Part III

Model

Chapter 7

Adaptive Lasso

For the prediction of credit loss rates in top-down stress testing, we propose the use of the Adaptive Lasso method. This method allows us to construct a flexible estimation procedure, which selects explanatory variables automatically from an initial set of predictors, where the number of predictors may even exceed the number of observations. The complexity and size of the set of initial predictors depends on the specification of the stress scenario that serves as input, and (non-linear) transformations that are applied to these macro-economic variables. The construction of this set of initial predictors and the application of Adaptive Lasso to the forecasting of NCO rates will be discussed in Chapter 8.

In this chapter we present the main theory behind Adaptive Lasso. We motivate the use of Adaptive Lasso by proving that it has many enticing finite-sample, asymptotic and computational properties. It has been shown that, in many cases, the Lasso method produces biased results, and only satisfies favorable convergence properties under stringent conditions. This is surely an undesirable result. Therefore, we follow Zou (2006) [16], and propose the Adaptive Lasso procedure instead. Adaptive Lasso retains the attractive features of the Lasso method, but it has the added bonus that it satisfies certain desirable convergence properties.

The remainder of this chapter is organized as follows. We start off with a short introduction to data-driven model discovery for NCO rates in Section 7.1. We continue with the introduction of Lasso as a special type of penalized regression, and develop some intuition as to how it works in Section 7.2. We proceed with the derivation of sparsity and continuity properties in Section 7.3. Subsequently, in Sections 7.5 and 7.4 we derive properties regarding unicity and convergence of Lasso and Adaptive Lasso. We present an application of the Lasso to variable screening in Section 7.6.1. This application will be used for a non-linear model which will be introduced in the next chapter. Lastly, in Section 7.6.2 we discuss the extension of Adaptive Lasso to panel data models. We conclude with a summary of the main results of this chapter in Section 7.7.

7.1 Data-driven Model Discovery

In Chapter 3 we suggested to separate the modeling of NCO rates of an individual bank, conditional on a macro-economic scenario into three distinct problems. First, we model the relationship between average NCO rates and macro-economic developments. The main model that we consider is a simple linear model of the form,

$$\mathrm{NCO}_t^{\mathrm{av}} = X_t\beta + \epsilon_t, \tag{7.1}$$

where ϵ_t are i.i.d. (independent identically distributed) random variables with mean zero. We let X be a large $(n \times p)$ design matrix consisting of potentially relevant predictors. In the second step we include bank-specific variables in a panel data model. The predictions of the models for single loan categories are aggregated to obtain forecasts for the NCO rates of an individual bank. A detailed discussion of these models is given in Chapter 8.

The modeling of credit loss rates conditional on the macro-economy can be approached from a statistical point of view or from a machine learning perspective. In the last case, we consider how we can discover, implement, and evaluate hypotheses generated by an automated selection procedure for modeling purposes and prediction. In this chapter we show that, Adaptive Lasso can be used to discover which transformations of macro-economic variables describe the regularities in the behavior of NCO rates in historical data best. Moreover, the coefficients of such a model can be estimated simultaneously.

The Adaptive Lasso procedure is an example of a supervised learning algorithm. Doornik & Hendry (2015) argue that automated selection procedures, such as Adaptive Lasso, can be productive when done correctly, especially in cases where the correct model cannot be known in advance and must therefore be discovered by a data-driven method. This is precisely the case for a model in top-down stress testing. Note that it remains to be seen whether there even exists a correct model for the relation between macro-economic developments and credit losses. In order to avoid misspecification or spurious regression when considering a large quantity of explanatory variables, the following subproblems must be resolved [51].

i Initial formulation: All potentially relevant variables and transformations thereof should be included in a set of candidate predictors, so that the true model can be discovered from the initial formulation.

ii Selection problem: Variables that are relevant need to be retained and effects that do not matter need to be discarded. This requires an appropriate selection procedure.

iii Computational problem: One needs a design of an approach which can handle the selection problem for the hefty amount of candidate variables in the initial formulation.

iv Evaluation problem: This concluding step consists of several methods. It should be checked whether the model is well-specified, which can be done by testing the residuals and assumptions of the model. Moreover, the discovered model should be economically reasonable and interpretable. In our case especially, the procedure should yield a high prediction accuracy on a test set.

In this chapter we focus on the issue of selecting a suitable design matrix X. To that end, we introduce Adaptive Lasso as an appropriate method for the selection and computational problems in data-driven model discovery. We develop the theory of the NCO rate models and the set-up of the top-down stress testing method and consider the implementation of the first three steps for our top-down stress testing method in Chapter 8. The evaluation problem will be addressed in Chapter 9.

7.2 Penalized Regression and the Lasso

Optimal model selection is still an unresolved problem in statistics, and there exist several commonly used approaches for this task. A very popular method is to use expert opinions or findings in previous theoretical or empirical literature as a starting point for a model. Variables are included or excluded from the model based on economic and statistical significance. There exist many methods to check whether the model fits the data well, and out-of-sample analysis can be employed to test the model's predictive strength. On the other hand, we have *automated* selection procedures, which generate hypotheses about the model and are tested based on out-of-sample prediction accuracy. Methods in this last category are an increasingly popular approach to model selection. Examples of automated methods include the use of unsupervised and supervised learning algorithms such as neural networks and penalized regression.

Especially when there are many potential predictors, it can be challenging to select an adequate model using traditional methods. The main issues are concerned with determining which variables and interactions to include in the model. The difficulty arises from the fact that as the number of predictors in the model increases, the bias decreases but its variance increases. If a model is used to forecast, we wish to minimize the overall prediction error, which is often referred to as optimizing the *bias-variance trade-off*[1].

[1]For the interested reader we discuss the bias-variance trade-off further in Appendix B.

Intuitively, the bias-variance trade-off can be interpreted as follows. As the complexity of a model increases, for instance by including more variables, complex structures in the data can be picked up more easily, which reduces the bias. On the other hand, coefficient estimates suffer from an increased variance due to the inclusion of more variables, and therefore the variance of the resulting model increases. This problem is often referred to as *over-fitting*. The variance induced by a high number of variables that are included in the model, may be controlled by introducing *regularization* on the coefficients.

The Lasso method is a *regularization* technique, which is the name for a set of methods that can be used to resolve an ill-posed problem or to prevent over-fitting. A popular regularization technique is penalized regression, where a penalty is added to an objective function. For example, the objective in linear regression is the minimization of the sum of squared residuals, whereas the objective in Maximum Likelihood Estimation (MLE) is the maximization of the likelihood function[2]. The penalty typically increases as more variables are included in the model, thus prioritizing more relevant predictors over lesser important ones.

A popular method to estimate the coefficients β for a linear model of the form,

$$Y = X\beta + \epsilon$$

is the renowned Ordinary Least Squares (OLS) method. It minimizes the squared residuals, $||Y - X\beta||_2^2$, and has a closed-form solution,

$$\hat{\beta} = \arg\min_{\beta} \frac{1}{2n} \sum_{i=1}^{n} (y_i - X_i\beta)^2 = (X^\top X)^{-1} X^\top Y.$$

When $p > n$, we have that $X^\top X$ is singular, and therefore a unique solution does not exist[3]. If however, we assume that the solution is *sparse* in the sense of Definition 7.1 below, then β has coefficients which have value zero, and it may still be possible to identify the correct model β. Under which conditions β is uniquely identifiable, will be addressed in Section 7.5.

Definition 7.1. (Sparsity) A model β of p coefficients is sparse when the set of active predictors $S = \{j : \beta_j \neq 0,\, j = 1, \ldots, p\}$ is less than p (i.e. $|S| < p$).

For the modeling of complex structures such as the impact of the macro-economy on credit losses, a true model might not exist or be too complex to estimate when only a limited amount of data is available. When the number of observations simply

[2] The likelihood function is explained in Appendix B.

[3] See Appendix B for a derivation and a more detailed discussion of OLS solutions and their assumptions and limitations.

cannot be increased, no model can be adequately estimated without the sparsity assumption. Moreover, if a top-down model is used in the context of stress testing, a parsimonious model which can easily be interpreted is even a requirement. In what follows, we therefore assume that the true model β^0 is sparse.

A good and sparse estimator for the coefficient parameter β can be obtained by best subset selection. Such a model can be identified by adding a penalty based on the cardinality of the active set. This penalty is also referred to as the $\mathcal{L}^0$ penalty, and the resulting estimator is given by,

$$\hat{\beta}^{\mathrm{L0}} = \arg\min_{\beta} \frac{1}{2n} \sum_{i=1}^{n} (y_i - X_i\beta)^2 + \lambda|\{j : \beta_j \neq 0\}|, \tag{7.2}$$

where λ is a tuning parameter, controlling the sparsity or complexity of the model $\hat{\beta}^{\mathrm{L0}}$. This estimator is referred to as a least squares $\mathcal{L}^0$-penalized estimator. The number of models that are considered using this method, is of the order 2^p (since each variable is either included or excluded from the model). Unfortunately this results in a non-convex optimization problem, which is even an NP hard programming problem [34]. Therefore, solving equation 7.2 is computationally infeasible even for moderate n and p. Note that model selection based on the well-known Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) criteria is a type of maximum likelihood $\mathcal{L}^0$ penalized regression [34].

Definition 7.2. (Penalized Regression) Let J be a penalty function and $f(\beta)$ the objective loss function for the model β. The sparse penalized estimator $\hat{\beta}$ is given by

$$\hat{\beta} = \arg\min_{\beta} \frac{1}{2n} f(\beta) + \sum_{j=1}^{p} J_{\lambda_j}(|\beta|), \tag{7.3}$$

where λ is a tuning or regularization parameter.

Note that when f is the negative log-likelihood, equation 7.3 corresponds to penalized maximum likelihood estimation. When f is the $\mathcal{L}^2$-loss or sum of squared residuals, equation 7.3 becomes penalized least squares regression. From here on, we shall focus on minimizing the squared residuals for linear models, so we consider $f(\beta) = ||Y - X\beta||_2$.

Although the $\mathcal{L}^0$-norm penalized regression is computationally infeasible, we can apply other penalty functions $J(\cdot)$ to the sum of squared residuals. There is variety of other penalties that can be used, most notably the $\mathcal{L}^1$ norm, better known as the Lasso penalty. Some other frequently used penalty functions are displayed in Table 7.1 below. In the next section we shall focus on establishing what a good penalty function is.

Table 7.1: Penalty Functions

Name	Function $J_{\lambda,\alpha}(\beta)$
Lasso	$\lambda\|\|\beta\|\|_1 = \lambda \sum_{j=1}^{p} \|\beta_j\|$
Ridge	$\alpha\|\|\beta\|\|_2 = \alpha \sum_{j=1}^{p} \beta_j^2$
Elastic Net	$\lambda\|\|\beta\|\|_1 + \alpha\|\|\beta\|\|_2 = \sum_{j=1}^{p} \left(\lambda\|\beta_j\| + \alpha\beta_j^2\right)$
Bridge	$\lambda\|\|\beta\|\|_q = \sum_{j=1}^{p} \|\beta_j\|^q$, $q > 0$
SCAD	$\begin{cases} \lambda\|\|\beta\|\|_1 & \text{if } \|\|\beta\|\|_1 < \lambda \\ -\frac{\|\|\beta\|\|_1^2 - 2a\lambda\|\|\beta\|\|_1 + \lambda^2}{2(a-1)} & \text{if } \|\|\beta\|\|_1 > a\lambda \\ \frac{(a+1)\lambda^2}{2} & \text{if } \lambda < \|\|\beta\|\|_1 \leq a\lambda \end{cases}$

Recently, as the amount of available data has risen dramatically, the Lasso penalty has become an increasingly popular topic of interest in statistics, automated feature selection methods and many other applications. The Lasso penalizes the absolute size of coefficients and in the process shrinks some of the model's coefficients to zero, which results in a parsimonious model. For the selection problem that we encounter in top-down stress testing, this is an attractive property. A linear model with few predictors can easily be evaluated and interpreted from an economic viewpoint. Another great advantage of the Lasso estimator, compared to non-penalized estimators, is that models with more features than observations, (i.e. when $p > n$) can be estimated. In this case, the solution is sparse.

Definition 7.3. (Lasso estimator) A Lasso estimator $\hat{\beta}^{\textbf{lasso}}$ is a solution to the following minimization problem

$$\arg\min_{\beta} \frac{1}{2n} \sum_{i=1}^{n} (y_i - X_i\beta)^2 \quad \text{subject to} \quad \sum_{j=1}^{p} |\beta_j| \leq t. \tag{7.4}$$

Where the bound t is a tuning parameter controlling the sparsity of the design matrix.

Lemma 7.4 shows that the Lasso estimator in Definition 7.3 is equivalent to the penalized regression with the Lasso penalty that we familiarized ourselves with in Definition 7.2. That is, we show that the constrained and unconstrained minimization problems are equivalent. Note that when $t < 0$ the solution to equation 7.4 does not exist, in which case we say that the minimization problem is *infeasible*.

Lemma 7.4. *(Lagrange Dual) Assume that the minimization problem in definition 7.3 is strictly feasible, then its solution is equivalent to,*

$$\arg\min_{\beta} \frac{1}{2n} \sum_{i=1}^{n} (y_i - X_i\beta)^2 + \lambda||\beta||_1. \tag{7.5}$$

Proof. For the proof we refer to Appendix C. □

Since $\sum_{j=1}^{p} |\beta_j|$ is the $\mathcal{L}^1$-norm of the parameter vector, Lasso regression is also often referred to as $\mathcal{L}^1$-penalized regression.

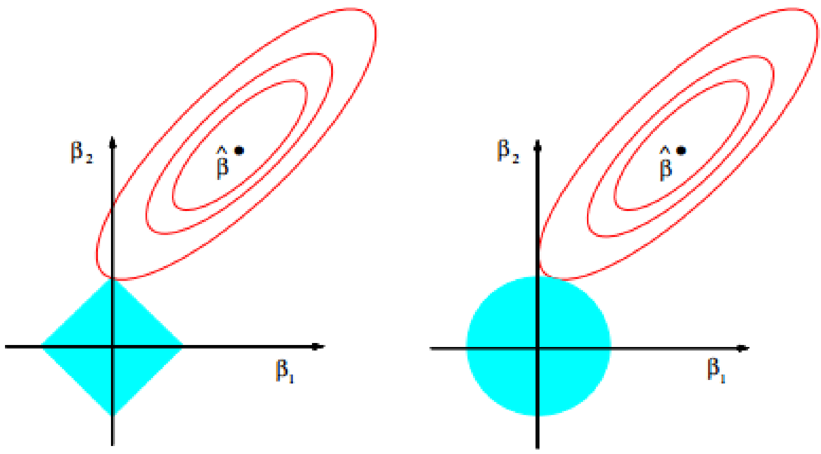


Figure 7.1: Estimation for lasso and ridge regression [4]

Figure 7.1 illustrates the intuition behind Lasso and ridge regression in the case of two potential predictors. The constrained regions are given by $|\beta_1| + |\beta_2| \leq t$ and $\beta_1^2 + \beta_2^2 \leq t^2$ for the lasso and ridge penalty[5] respectively, and are represented by the blue areas in the figure. The optimal solution to 7.3 lies within the blue constraint region. The OLS solution is displayed by $\hat{\beta}$ and the ellipses represent the sums of squared residuals for values of β.

Proposition 7.5 states that when the OLS solution lies outside of the constrained region, the Lasso solutions are found at the intersection of the shaded areas with the ellipses. If $\hat{\beta}$ were to lie in the shaded areas, then the solution to the penalized regression and OLS regression would be the same. Intuitively, the angular shape of the Lasso constrained region renders sparse results, whereas the smooth shape of the ridge constraint region does not. We remark that in Figure 7.1, estimates for large values of β_1 and β_2 are also shrunk because they have to be within the constraint region. In the following section we will proof that penalized regression results in a sparse estimator under the condition that its penalty function is non-differentiable at the origin.

Proposition 7.5 (Lasso Boundary Solution)**.** *If*

$$t < t_0 := \min_{\beta}\{||\beta||_1 : X^\top X\beta = X^\top Y\}, \tag{7.6}$$

[4]Image was taken from The Elements of Statistical Learning [19], figure 3.11.
[5]Ridge regression can be similarly defined in its primal and dual problem as Lasso regression.

then for a solution $\hat{\beta}$ *it holds that* $\sum_{j=1}^{p} |\beta_j| = t$. *When* $t \geq t_0$ *then some OLS estimate is the solution.*

Proof. For the proof we refer to Appendix F. $\square$

7.3 Sparsity and Continuity in Penalized Regression

Although the focus of this report is on the Adaptive Lasso penalty, we look at a broader range of penalty functions in this section to motivate its use. In the previous section we have discussed the Lasso as a special case of penalized regression. A sensible follow-up question is why (Adaptive) Lasso is a suitable penalty function. Fan & Li (2001) conjecture that a good penalty function possesses the three properties listed below [9]. Their definition of a good penalty function has been widely used, for instance in [16], and [19].

1. Approximate Unbiasedness: Suppose the true unknown parameter β_j^0, $j \in \{1, \ldots, p\}$ for a true predictor is large. The resulting penalized regression estimator $\hat{\beta}_j^{\text{pen}}$ should be nearly unbiased for such a large parameter, in order to avoid extra modeling bias. This means that if $\hat{\beta}_i^{\text{pen}}$ is estimated on data generated from the linear model $X\beta^0 + \epsilon^0$, it should hold that,
$$\mathbb{E}_{\beta^0}[\hat{\beta}_j^{\text{pen}} - \beta_j^0] \approx 0, \tag{7.7}$$
for large β_j^0. So that $X\beta_0 \approx X\hat{\beta}^{\text{pen}}$.
2. Sparsity: The penalty function should result in an estimator that sets small estimates of coefficients to zero. Let $\hat{S} = \{j : \hat{\beta}_j^{\text{pen}} \neq 0\}$, then sparsity implies that $|\hat{S}| < p$. This naturally reduces model complexity, which typically has a positive effect on our ability to interpret the resulting model.
3. Continuity: The estimator $\hat{\beta}^{\text{pen}}$ should be continuous for the input data to prevent modeling instability.

The remainder of this section is dedicated to deriving and understanding the conditions on $J(\cdot)$, which ensure that the resulting estimator of the penalized regression satisfies the above three properties. The last two conditions are outlined in Theorem 7.6.

Theorem 7.6. *A non-negative, non-decreasing, and convex penalty function that is non-differentiable at the origin and continuous on* $(0, \infty)$ *results in a sparse and continuous estimator.*

Proof. An outline of the proof is given in appendix F. Note that, the result also holds for non-convex penalty functions. The proof of this results can be found in [17]. □

Corollary 7.7. *The Lasso penalty function gives sparse and continuous solutions.*

Proof. The Lasso penalty is non-negative, non-decreasing, and convex[6]. Theorem 7.6 states that if the penalty function is non-differentiable at zero, it should result in a sparse solution. Since the Lasso penalty is non-differentiable at the origin and it is continuous on $(0, \infty)$, it produces sparse and continuous results. □

Theorem 7.8. *A sufficient condition for an approximately unbiased estimator is that the penalty function is bounded by a constant.*

Proof. We refer to Fan and Li (2001) [9] for the proof. □

The SCAD penalty is the only function in Table 7.1 that is bounded by a constant, and it can be shown that it produces an approximately unbiased estimator [17]. Although the Lasso method is a procedure that gives sparse and continuous results, the bias of the method is generally thought to be a problem.

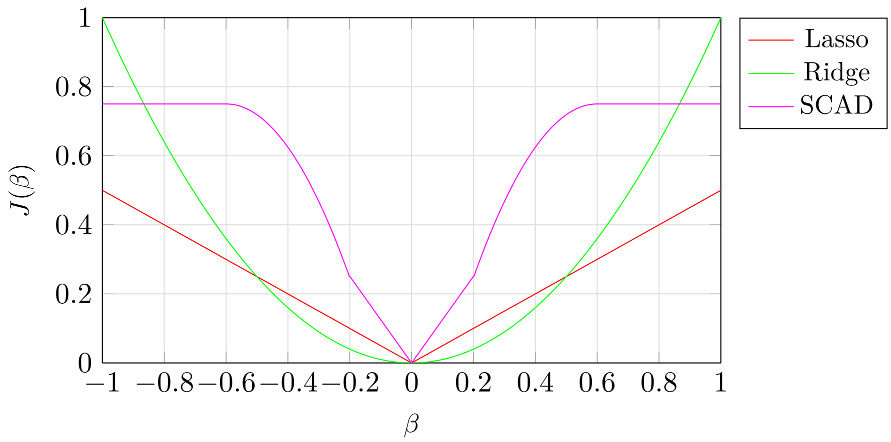


Figure 7.2: Penalty functions for $\lambda = 0.5$, $\alpha = 0.5$ and $a = 5$.

Theorem 7.6 indicates that a suitable penalty function is typically continuous and non-decreasing on $(0, \infty)$, singular at the origin, and bounded by a constant. In figure 7.2 the penalty some of the functions in Table 7.1 are displayed. This figure shows that only the SCAD penalty is bounded by a constant. However, it is also clear from Figure 7.2 that the SCAD penalty function is not convex. It is surely not as computationally efficient as the Lasso. Although the Lasso penalty is not

[6]Convexity of the Lasso penalty will be shown in Lemma 7.15 in Section 7.5.

bounded by a constant, *Adaptive* Lasso, which will be introduced in Section 7.4, does give approximately unbiased estimates for the coefficients.

In what follows, we discuss the approximate unbiasedness of penalized estimators. We first derive some intermediate results, and show that the minimization problem in Definition 7.2 for the $\mathcal{L}^2$ loss function can be written in a more convenient form.

Lemma 7.9. *Let $\hat{\beta}$ be a minimizer of $\sum_{i=1}^{n}(y_i - X_i\beta)^2$, then minimizing*

$$\frac{1}{2n}\sum_{i=1}^{n}(y_i - X_i\beta)^2 + \sum_{j=1}^{p} J_{\lambda_j}(\beta_j), \tag{7.8}$$

is equivalent to minimizing

$$Q(\beta) = \frac{1}{2n}(\beta - \hat{\beta})^\top X^\top X(\beta - \hat{\beta}) + \sum_{j=1}^{p} J_{\lambda_j}(\beta_j), \tag{7.9}$$

with respect to β.

Proof. For the proof of this lemma we refer to Appendix F. □

To better understand the bias problem, we look at *thresholding* rules, which are methods to iteratively update the coefficients for each coordinate of the penalized regression estimator. To that end, we first need the intermediate result in Lemma 7.10.

Lemma 7.10. *Let $f(x) = g(x) + \sum_{j=1}^{p} h_j(x_j)$, where $g(x)$ is convex and continuously differentiable, and $h_j(x)$ are convex functions. Then*

$$f(x + ce_j) \geq f(x), \quad \text{for all } c, j \Leftrightarrow f(x) = \min_{\beta \in \mathbb{R}^p} f(\beta), \tag{7.10}$$

where $e_j \in \mathbb{R}^p$ is the jth standard basis vector.

Proof. For the proof we refer to Appendix F. □

By Lemma 7.10, the global minimum of the constrained Lasso problem is found if and only if it is optimized along each axis. In the next chapter a cyclical coordinate descent algorithm will be introduced, that solves the penalized regression problem for convex penalty functions, by optimizing along each direction iteratively until the process converges. The convergence properties of convex optimization, will be stated there. But here we use this property to derive so-called thresholding rules and shrinkage factors for different types of penalties. We derive these to illustrate the bias of ridge and Lasso estimates.

To simplify the results, we only show the derivation of the thresholding rules for the case that $\hat{\mathcal{I}} = X^\top X$ is the identity matrix (X is orthonormal). The following results are only used to demonstrate the behavior of thresholding and shrinkage in penalized regression. First we look at the effect of optimizing the ridge penalized regression in one direction, keeping all other coordinates fixed. To that end, we fix β_k for $k \neq j$, and let

$$\beta_j^{\text{ridge}} = \arg\min_{\beta_j} \frac{1}{2n} \sum_{i=1}^{n} (y_i - X_i\beta)^2 + \sum_{j=1}^{p} \beta_j^2. \tag{7.11}$$

Setting the derivative of $Q(\beta)$ in Lemma 7.9 to zero, we derive the *shrinkage factor* for ridge penalized regression. We write,

$$\begin{aligned}
0 = \frac{\partial Q(\beta^{\text{ridge}})}{\partial \beta_j^{\text{ridge}}} &= \sum_{k=1}^{p} \hat{\mathcal{I}}_{jk}(\beta_k^{\text{ridge}} - \hat{\beta}_k) + 2\lambda\,\beta_j \\
\implies (\hat{\beta}_j^{\text{ridge}} - \hat{\beta}_j) + 2\lambda\,\beta_j^{\text{ridge}} &= 0 \\
\implies \hat{\beta}_j^{\text{ridge}} &= \frac{1}{1+2\lambda}\hat{\beta}.
\end{aligned}$$

From this, we conclude that optimizing the ridge estimator in direction j results in a shrinkage of the OLS estimate by a factor $\frac{1}{1+2\lambda}$. The shrinkage effect as a function of β is displayed in Figure 7.3c.

In a similar way we derive the *soft thresholding* rule for the Lasso penalty. We write,

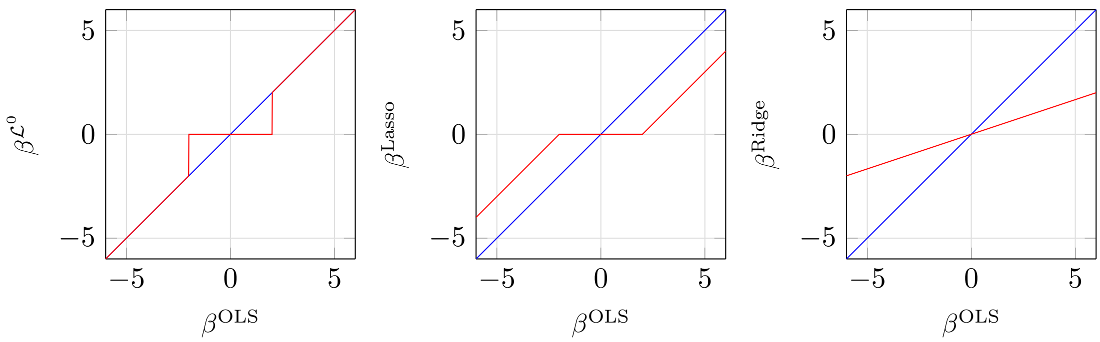
$$\begin{aligned}
0 = \frac{\partial Q(\beta)}{\partial \beta_j} &= \sum_{k=1}^{p} \hat{\mathcal{I}}_{jk}(\beta_k - \hat{\beta}_k) + \lambda\,\text{sgn}(\beta_j) \\
\implies & (\hat{\beta}_j^{\text{lasso}} - \hat{\beta}_j) + \lambda\,\text{sgn}(\hat{\beta}_j^{\text{lasso}}) = 0 \\
\implies & \text{sgn}(\hat{\beta}_j^{\text{lasso}}) \left(|\hat{\beta}_j^{\text{lasso}}| + \lambda\right) - \hat{\beta}_j = 0 \\
\implies & \hat{\beta}_j^{\text{lasso}} = \begin{cases} 0 & \text{if } |\hat{\beta}_j| \leq \lambda \\ \left(1 - \frac{\lambda}{|\hat{\beta}_j|}\right)\hat{\beta}_j & \text{otherwise} \end{cases} \\
\implies & \hat{\beta}_j^{\text{lasso}} = \text{sgn}(\hat{\beta}_j)(|\hat{\beta}_j| - \lambda)^+,
\end{aligned}$$

where we set the derivative of $Q(\beta)$ to zero component-wise in the second implication. The transformation by the soft-thresholding rule of the initial estimate $\hat{\beta}$, which can be obtained by the OLS method for orthonormal design matrix X, is displayed in Figure 7.3b.

Best subset selection, or $\mathcal{L}^0$ penalized regression, which we mentioned in equation 7.2 for orthonormal designs coincides with the so-called *hard thresholding* rule, which is given by [9]:

$$\hat{\beta}_j^{\mathrm{L0}} = \hat{\beta}\mathbb{1}_{\{|\hat{\beta}_j| > \lambda\}}. \tag{7.12}$$

In Figure 7.3a the effect of this transformations of the initial $\hat{\beta}$ estimate is displayed.



(a) Hard Thresholding (b) Soft Thresholding (c) Shrinkage

Figure 7.3: Transformations of the unbiased estimator $\hat{\beta}^{\mathrm{OLS}}$, for the hard-thresholding rule, the soft-thresholding rule, and ridge shrinkage. These thresholding rules are implied by the $\mathcal{L}^0$, Lasso, and ridge penalty respectively.

We see that, although Lasso renders continuous and sparse results, this comes at a price of shifting the OLS estimator by a constant λ [9]. The introduction of such a bias in the orthonormal case, suggests that the Lasso procedure can give biased estimates. A proof of the bias of Lasso is given in Zou (2006) [16]. In the following section we introduce Adaptive Lasso, which retains the favorable properties but adaptively chooses weights for coefficients to reduce bias.

7.4 The Oracle Properties of Adaptive Lasso

The literature on Lasso and penalized regression is vast. Besides the sparsity and continuity properties that we have discussed so far, the asymptotic properties of automated selection procedures are also a flourishing topic of research. A substantial amount of effort is being put forward to derive robust measures for the errors of an Adaptive Lasso estimator. The least squares $\mathcal{L}^0$ estimator $\hat{\beta}^{\mathrm{L0}}$ in equation 7.2 is thought to be a good estimator because it enjoys the *oracle* properties [34]. In this section, we give a precise definition of such an oracle estimator, and show that Adaptive Lasso also possesses these desirable properties. The oracle properties ensure that variable selection is consistent, and that the estimator $\hat{\beta}$ asymptotically converges to the true coefficients, at the same rate as if the true parameters were known in advance.

Before we proceed, we establish some useful notation for asymptotic results. Recall that S^0 is the active set of the true underlying model. In the asymptotic setting we consider the Lasso estimator

$$\hat{\beta}^{(n)} = \arg\min_{\beta} \frac{1}{2n}||Y^{(n)} - X^{(n)}\beta||_2^2 + \lambda_n||\beta||_1, \tag{7.13}$$

and let $n \to \infty$. Let δ be the procedure used to obtain the estimator $\hat{\beta}^{(n)}(\delta)$. We denote the active set of variables of this estimator by $S^{(n)} = \{j = 1, \ldots, p : \hat{\beta}_j^{(n)}(\delta) \neq 0\}$, and let $\beta^0_{S^0}$ denote the nonzero coefficients of β^0. Similarly, $\hat{\beta}^{(n)}_{S^{(n)}}(\delta)$ represents the nonzero coefficients of $\hat{\beta}^{(n)}(\delta)$. Lastly, we let Σ^0 be the covariance matrix for the true subset model.

Definition 7.11. (Oracle Properties) Let $\hat{S}^{(n)}$ and $(\hat{\beta}(\delta)$ be the active set and the model estimate obtained by a procedure δ on n observations. We say that a procedure δ satisfies the oracle properties when:

i It selects variables consistently,

$$\lim_{n\to\infty} \mathbb{P}(S^{(n)} = S^0) = 1. \tag{7.14}$$

ii It has the asymptotic distribution as if the true active set was known in advance,

$$\sqrt{n}\left(\hat{\beta}^{(n)}(\delta)_{S^0} - \beta^0_{S^0}\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma^0\right). \tag{7.15}$$

The first oracle property ensures that precisely those predictors in the true model are selected as n goes to infinity. The second oracle property tells us that asymptotically, the procedure performs as well as if the variables in the active set were known in advance.

A large portion of the research dedicated to Lasso techniques is focused on deriving necessary and sufficient conditions for the Lasso procedure and variants thereof to have the oracle properties. Whether the Lasso procedure possesses the oracle properties has been a topic of debate, see for instance [18] [15] [28] [32]. It turns out that it satisfies the oracle properties under non-trivial conditions on the underlying model. A thorough overview of the conditions that are used to prove oracle results for Lasso can be found in S. van de Geer & P. Bühlmann (2009) [18]. Zou(2006) gives scenarios for which the Lasso selection cannot be consistent [16]. In the same paper he proposes the use of the *Adaptive* Lasso.

In the Lasso procedure all coefficients are penalized by the same tuning parameter λ and thus receive the same weight, independent of their size. As a result, large coefficients contribute relatively more to the Lasso penalty, which can lead to biased results [16]. Note that our observation from Figure 7.1 that large coefficients are shrunk towards zero corresponds to this result. The Adaptive Lasso in Definition 7.12 allows each coefficient to have a different, adaptively chosen weight.

Definition 7.12. (Adaptive Lasso) Suppose $\hat{\beta}^{\text{init}}$ is a root-n consistent estimator[7] for β^0. Then the Adaptive Lasso estimator $\beta^{(n)}$ is given by

$$\hat{\beta}^{(n)} = \arg\min_\beta ||Y^{(n)} - X^{(n)}\beta||_2^2 + \lambda_n \sum_{j=1}^{p} \hat{w}_j |\beta_j|, \tag{7.16}$$

where $\hat{w} = |\hat{\beta}^{\text{init}}|^{-\gamma}$ an adaptive weight vector.

The parameter γ in the definition of the weights for Adaptive Lasso, can be chosen so that the prediction error of the procedure is minimized. In Section 8.1.3 in the next chapter, we show how this optimal value of γ can be computed.

It can still be argued that a $\mathcal{L}^0$ penalty as in equation 7.2 is more natural than the $\mathcal{L}^1$ penalty, since it directly penalizes the inclusion of extra parameters. The main advantage of a $\mathcal{L}^0$ penalty is that it satisfies the oracle properties in Definition 7.11. It turns out that the Adaptive Lasso method enjoys the oracle properties under far less stringent conditions than Lasso. Recall that σ^2 denotes the variance of the i.i.d. random variables $\epsilon = (\epsilon_1, \ldots, \epsilon_n)$ with mean zero in the linear model,

$$y_i = X_i\beta^0 + \epsilon_i.$$

We assume that $\frac{1}{n}X^{(n)\top}X^{(n)} \to C$, where C is a positive definite matrix. Now without loss of generality, we let $S^0 = \{1, \ldots, p_0\}$, where $1, \ldots, p_0$ are the true predictors. We rewrite,

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}, \tag{7.17}$$

where C_{11} is a $p_0 \times p_0$ matrix representing the covariance matrix of the true active set of predictor variables.

Theorem 7.13. *Let $\lambda_n/\sqrt{n} \to 0$ and $\lambda_n n^{(\gamma-1)/2} \to \infty$, then the adaptive lasso estimates in Definition 7.12 satisfy the oracle properties,*

i $\lim_{n\to\infty} \mathbb{P}(S^{(n)} = S^0) = 1$,

ii $\sqrt{n}\left(\hat{\beta}^{(n)}_{S_0}(\delta) - \beta^0_{S_0}\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 \times C_{11}^{-1}\right)$

Proof. For the proof of this theorem we refer to Zou (2006) [16]. In Appendix F we give an outline of the proof. We note that the consistency in variable selection also holds for Lasso and that the results for Adaptive Lasso hold for fixed p. Huang et

[7] A $\sqrt{(n)}$-consistent estimator $\hat{\beta}$ satisfies $\hat{\beta}^{(n)} - \beta^0 = O_p(n)$.

al. (2008) show that the results for Adaptive Lasso hold for the high-dimensional case, where $p \gg n$ under some extra conditions [14]. $\square$

We now know that the adaptive lasso is an asymptotically consistent selection and estimation procedure for some $\sqrt{n}$-consistent initial estimator $\hat{\beta}^{\text{init}}$. Unfortunately, the choice of $\hat{\beta}^{\text{init}}$ is still somewhat of an open problem. Zou (2006) suggests that the ridge regression estimator can be an adequate choice [16] in the $p > n$ case. The use of Lasso estimators for $\hat{\beta}^{\text{init}}$ can also be found in literature and is advocated by Bühlmann and Geer for high-dimensional settings in [26]. For this research project we considered $\beta^{textinit} = \beta^{textlasso}$.

The Adaptive Lasso estimator satisfies the same properties with regard to sparsity and continuity as does the Lasso [16]. The Adaptive Lasso penalty is non-decreasing, non-negative, non-differentiable at zero, and also convex. Hence by Theorem 7.6 it produces sparse and continuous results. The Adaptive Lasso estimates are approximately unbiased by Theorem 7.13.

To exemplify this, we demonstrate the behavior of the thresholding rule for the Adaptive Lasso in the case that X is orthonormal, as we did for Lasso in the previous section. This exercise is intended to illustrate the difference with the Lasso and the influence of the parameter γ in the weight function. Applying the same approach as in section 7.3, we can derive the associated thresholding function,

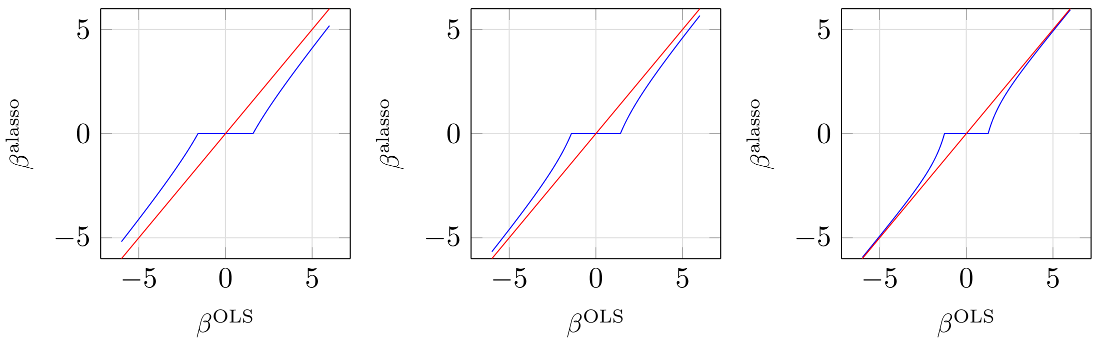
$$\hat{\beta}_j^{\text{alasso}} = \operatorname{sgn}(\hat{\beta}_j)\left(|\hat{\beta}_j| - \frac{\lambda}{|\hat{\beta}_j|^\gamma}\right)^+, \tag{7.18}$$

where $\hat{\beta}$ is a consistent and unbiased estimator, for instance βOLS. In Figure **??** below, we observe that the thresholding function of the Adaptive Lasso penalty renders approximately unbiased results, because the difference $\hat{\beta}_j^{\text{alasso}}$ and $\hat{\beta}_j$ goes to zero as $\beta \to \infty$. Convergence of the threshold function to the OLS solution is faster when γ is larger.

7.5 Existence and Uniqueness of Lasso Solutions

In this section we give explicit proofs for the existence and uniqueness of Lasso estimators. All results can easily be adapted for Adaptive Lasso, but for simplicity in notation we show the results for Lasso instead. We use theory from convex optimization to prove the existence of the Lasso estimator and derive conditions under which it is unique. The Lasso problem belongs to the category of convex optimization problems without equality constraints. It can be shown[8] that the

[8]See Appendix C for an introduction to convex optimization theory and the derivation of the KKT conditions.



(a) $\gamma = 0.5$ (b) $\gamma = 1$ (c) $\gamma = 2$

Figure 7.4: Transformations of the OLS estimate for the Adaptive Lasso thresholding function.

solution to the constrained Lasso problem in equation 7.4 satisfies the Karush-Kuhn-Tucker (KKT) conditions in Definition 7.14 below. Moreover, from optimization theory we know that a solution $\hat{\beta}$ to the Lasso problem in equation 7.4 is optimal if and only if it satisfies these KKT conditions.

Definition 7.14. (KKT conditions for Lasso) The KKT conditions for the Lasso problem are given by

i Stationarity: $X^\top(Y - X\hat{\beta}) = \lambda s$, where

$$s_j \in \begin{cases} \{1\} & \hat{\beta}_j > 0, \\ [-1, 1] & \hat{\beta}_j = 0, \\ \{-1\} & \hat{\beta}_j < 0. \end{cases}$$

ii Complementary Slackness: $\lambda[\sum_{j=1}^{p} |\beta_j| - t] = 0$

iii Primal Feasibility: $\sum_{j=1}^{p} |\beta_j| \leq t$

iv Dual Feasibility: $\lambda \geq 0$

The KKT conditions play a crucial part in some of the theorems that are discussed in this section. Another important concept is convexity. In Lemma 7.15 we specifically state some results with respect to norms and convexity, which are essential to the proofs of the theorems in the remainder of this section.

Lemma 7.15. *(Convexity) We say that a function h is convex when*

$$h(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha h(x_1) + (1 - \alpha)h(x_2), \tag{7.19}$$

for $\alpha \in [0, 1]$ and strictly convex if 7.19 has strict inequality for $\alpha \in (0, 1)$. The following statements hold:

i *A strictly convex function f has at most one minimization point.*

ii *A convex function f on a closed and bounded set $\mathcal{C}$ has at least one minimization point.*

iii *If h is a strictly convex function and f is a convex function, then $f + h$ is a strictly convex function.*

iv *The function $h(\beta) = ||\beta||_1$ is convex.*

v *The function $h(X\beta) = ||Y - X\beta||_2^2$ is strictly convex and the function $h(\beta) = ||Y - X\beta||_2^2$ convex.*

vi *If the columns of X are linearly independent $h(\beta) = ||Y - X\beta||_2^2$ is strictly convex in β.*

Proof. The proof of this lemma can be found in Appendix F. □

Theorem 7.16 (Existence)**.** *The solution to the Lasso problem,*

$$\arg\min_{\beta} \frac{1}{2n} \sum_{i=1}^{n} (y_i - X_i\beta)^2 + \lambda||\beta||_1. \tag{7.20}$$

is either unique or there exist infinitely many solutions. If the columns of X are linearly independent, then the solution is unique.

Proof. We know that $f(\beta) = ||Y - X\beta||_2^2$ is a convex function by lemma 7.15 (5). Since the set $\{\beta \in \mathbb{R}^p : ||\beta||_1 \leq t\}$ is closed and bounded, a solution to the lasso problem is guaranteed to exist by lemma 7.15(2). Moreover, if the columns of X are linearly independent $f(\beta)$ is strictly convex by 7.15(6), the objective in equation 7.20 is strictly convex by 7.15(3), and hence the solution is unique by 7.15(1).

We have established that there exists at least one solution. We show that if there exist more than one solution, than there exist uncountably many solutions. Let $f(\beta) = \frac{1}{2n}||Y - X\beta||_2^2 + \lambda||\beta||_1$, and suppose there exist two solutions $\hat{\beta}^{(1)} \neq \hat{\beta}^{(2)}$, then $f(\hat{\beta}^{(1)}) = f(\hat{\beta}^{(2)})$. Since the constraint region $||\beta||_1 \leq t$ is convex we have that $\alpha\hat{\beta}^{(1)} + (1-\alpha)\hat{\beta}^{(2)}$, for $0 < \alpha < 1$, is in the feasible set. By convexity of f we may write

$$f(\alpha\hat{\beta}^{(1)} + (1-\alpha)\hat{\beta}^{(2)}) \leq \alpha f(\hat{\beta}^{(1)}) + (1-\alpha)f(\hat{\beta}^{(2)}) = f(\hat{\beta}^{(1)}). \tag{7.21}$$

The inequality is actually an equality, because otherwise $f(\alpha\hat{\beta}^{(1)} + (1-\alpha)\hat{\beta}^{(2)}) < f(\hat{\beta}^{(1)})$ contradicts the fact that $f(\hat{\beta}^{(1)})$ is a minimal solution. It follows that $\alpha\hat{\beta}^{(1)} + (1-\alpha)\hat{\beta}^{(2)}$ is also a minimal solution. Since $\alpha \in [0, 1]$ there are uncountably many solutions. □

Theorem 7.16 shows that the Lasso estimator is guaranteed to exist. Moreover, if the columns of X are linearly independent then the solution is unique. This sufficient condition for uniqueness can be made less stringent, so that Lasso solutions for problems involving design matrices with more predictors than observations (i.e. $p > n$) can also be uniquely determined. Before we continue with the condition for uniqueness of the Lasso solution, we first establish another important result, which guarantees that the predictions that the model renders are unique in theorem 7.17.

Theorem 7.17 (Unique Predictions)**.** *The predictions of Lasso estimators are always unique.*

Proof. To proof the result we show that if the Lasso minimization problem has two solutions $\hat{\beta}^{(1)}$ and $\hat{\beta}^{(2)}$ then $X\hat{\beta}^{(1)} = X\hat{\beta}^{(2)}$. We argue by contradiction. Let

$$c^* = ||Y - X\hat{\beta}^{(1)}||_2^2 + \lambda||\hat{\beta}^{(1)}||_1 = ||Y - X\hat{\beta}^{(2)}||_2^2 + \lambda||\hat{\beta}^{(2)}||_1. \tag{7.22}$$

For any $0 < \alpha < 1$ we may write

$$||Y - X\left(\alpha\hat{\beta}^{(1)} + (1-\alpha)\hat{\beta}^{(2)}\right)||_2^2 + \lambda||\alpha\hat{\beta}^{(1)} + (1-\alpha)\hat{\beta}^{(2)}||_1 < \alpha c^* + (1-\alpha)c^* = c^*, \tag{7.23}$$

where the strict inequality follows from the fact that $||Y - X\beta||_2^2$ is a strictly convex function of $X\beta$, and $||\beta||_1$ a convex function. But then the solution $\alpha\hat{\beta}^{(1)} + (1-\alpha)\hat{\beta}^{(2)})$ gives a value less than c^*. This contradicts the fact that $\hat{\beta}^{(1)}$ and $\hat{\beta}^{(2)}$ are Lasso solutions. The inequality in equation 7.23 is an equality if and only if $X\hat{\beta}^{(1)} = X\hat{\beta}^{(2)}$. Hence the Lasso predictions are unique. □

Theorem 7.17 shows us that the Lasso produces unique predictions for any matrix X. However, we would like to know under what conditions the solution itself is unique. From a modeling perspective, uniqueness is a useful property, since there is no ambiguity concerning the model specification, which makes model interpretation easier. Moreover, it shows us that the Lasso procedure can distinguish between many subtly different models. Since we want to use Adaptive Lasso to simultaneously select and estimate a parsimonious model from a large set of candidate predictors, this is a very useful property. In Theorem 7.16 we proved that the uniqueness of the Lasso estimator is unique in the case when the columns of X are linearly independent. Surely, linear independence does not hold for the case $p > n$. In Theorem 7.18 below we show that uniqueness can be established under far less stringent conditions.

Theorem 7.18 (Uniqueness)**.** *Let $X \in \mathbb{R}^{n\times p}$ and $\lambda > 0$. Consider*

$$\underset{\beta\in\mathbb{R}^p}{\arg\min}\, ||Y - \beta X||_2^2 + \lambda||\beta||_1. \tag{7.24}$$

Define $E = \{j = \{1, \ldots, p\} : X^\top(Y - X\hat{\beta}) = \lambda\}$ *and let* X_E *denote the matrix of the columns of* X *that are in* E*. If* X_E *has full column rank then the solution* $\hat{\beta} \in \mathbb{R}^p$ *to 7.24 is unique and has at most* $\min(n, p)$ *nonzero components.*

Proof. Recall the stationarity condition of the KKT conditions, $X^\top(Y - X\hat{\beta}) = \lambda s$. Since $X\hat{\beta}$ is unique by theorem 7.17, it follows that s, and hence E is uniquely determined.

Note that for $j \notin E$ it holds that $|s_j| < 1$, which implies that $\hat{\beta}_j = 0$. Therefore, E contains the set of active variables, and $\hat{\beta}_{-E} = 0$. For any solution $\hat{\beta}$ we must have, again by the KKT conditions,

$$X_E^\top(Y - X_E\hat{\beta}) = \lambda s_E. \tag{7.25}$$

A particular solution of this system of linear equations is given by

$$\hat{\beta}_E = (X_E^\top X_E)^\dagger (X_E^\top Y - \lambda s_E), \tag{7.26}$$

where $\dagger$ denotes the Moore-Penrose pseudo-inverse. The general form of the solution is given by,

$$\hat{\beta}_E = (X_E^\top X_E)^\dagger (X_E^\top Y - \lambda s_E) + \eta, \tag{7.27}$$

where $\eta \in \ker(X_E^\top X_E) = \ker(X_E)$.

From here, it is straight-forward to see that the solution $\hat{\beta}$ is unique if and only if $\ker(X_E) = \{0\}$ (recall that X_E is uniquely determined). It follows that if X_E has full column rank then the lasso solution is unique and satisfies

$$\hat{\beta}_E = (X_E^\top X_E)^{-1} (X_E^\top Y - \lambda s_E), \qquad \hat{\beta}_{-E} = 0. \tag{7.28}$$

The fact that X_E has full column rank implies that $|E| \leq \min(n, p)$. □

Actually, it can be shown that except for $y \in \mathbb{R}^n$ in a set of measure zero it holds that $\forall j \in E$, $\hat{\beta}_j \neq 0$ [39]. It follows that $S = E$ and the conditions of theorem 7.18 are not only sufficient but also almost everywhere necessary. What remains is to consider conditions under which X_E has full column rank. This is trivially true in the case that X itself has full column rank. Theorem 7.21 below demonstrates conditions under which X_E has full column rank.

Definition 7.19. (Affine Span) The affine span of vectors $X_1, \ldots X_p$, is given by

$$\left\{ \alpha_1 X_1 + \ldots, \alpha_p X_p : \alpha \in \mathbb{R}^p, \sum_{j=1}^{p} \alpha_j = 1 \right\}. \tag{7.29}$$

Definition 7.20. (General Position) We say that $X_1, \dots, X_p \in \mathbb{R}^n$ are in *general position* when for all $k < \min(n, p)$, $i_1, \dots, i_{k+1} \in \{1, \dots, p\}$, and $\sigma_1, \dots, \sigma_{k+1} \in \{-1, 1\}$ the affine span of $\sigma_1 X_{i_1}, \dots \sigma_{k+1} X_{i_{k+1}}$ contains no elements of $\{\pm X_i : i \neq i_1, \dots, i_{k+1}\}$.

Theorem 7.21. *If the columns of X are in general position, then X_E has full column rank.*

Proof. For the proof we refer to Tibshirani (2013) [39]. □

The condition that the columns of X need to be in general position for uniqueness of the lasso solution, is very weak and can even hold for design matrices X with $p \gg n$. These are very fortunate results for the (Adaptive) Lasso method. In fact, the solutions are only non-unique when the columns are not in general position, which is the case when exact collinearity exists in relatively small sets of variables. Theorem 7.22 below gives a condition on the continuity of the distribution of X that also ensures unique solutions.

Theorem 7.22. *If the entries of X are drawn from a continuous probability distribution on $\mathbb{R}^{n \times p}$, with respect to the Lebesgue measure, then the solution $\hat{\beta} \in \mathbb{R}^p$ to 7.24 is unique.*

Proof. It can be shown that if the entries of X are drawn from a continuous distribution on $\mathbb{R}^{n \times p}$ then the columns are in general position, for a proof we refer to [39]. By Theorem 7.21 this implies that X_E has full column rank. Theorem 7.18 finishes the proof. □

7.6 Other Applications of Adaptive Lasso

In this section we describe two applications of Lasso that can be used for more advanced models. In Section 7.6.1 we demonstrate that Adaptive Lasso can be used as a variable screening method that retains substantially relevant predictors. The application of Adaptive Lasso to panel data models will be discussed in Section 7.6.2

7.6.1 Variable Screening

In the previous sections we have demonstrated that the Adaptive Lasso method has many desirable properties for variable selection and modeling. In this section we discuss the option of using (Adaptive) Lasso for *variable screening*. This application refers to a method that reduces the dimensionality of the design matrix. In the following chapter we will use this option for one of our benchmark models.

Variable screening is particularly useful for high-dimensional modeling, where typically $p > n$. Most statistical methods cannot produce estimates under such circumstances. A reduced design matrix may also allow the option of using other variable selection methods, which would otherwise be computationally infeasible. For large design matrices where $p \gg n$ it can therefore be useful to perform variable screening on the design matrix. To that end, we want to find predictors whose coefficients $|\beta_j|$ are *substantially relevant*.

Definition 7.23. (Substantially Relevant Predictors) Recall that the columns of X are standardized, so that they have equal mean and variance. The set of substantially relevant predictors in the true model is given by,

$$\mathcal{S}_0^C = \{j : |\beta_j^0| \geq C, \, j = 1, \dots, p\}, \tag{7.30}$$

for some $C > 0$.

Let $\hat{\beta}(\lambda^*)$ be the Adaptive Lasso estimator for the optimal value of the tuning parameter. How this value can be obtained computationally will be discussed in 8.1.3 in the next chapter. It can be shown that $||\hat{\beta}(\lambda*) - \beta^0||_q \to 0$ in probability by using Theorem 7.13. It can then be derived that for all fixed $0 < C < \infty$ it holds that

$$\mathbb{P}(\hat{S}(\lambda^*) \supseteq S_0^C) \to 1 \qquad \text{as } n \to \infty. \tag{7.31}$$

We refer to Bühlman and Geer (2011) [25] for a thorough derivation of these results. Lemma 7.24 below implies that if we perform (Adaptive) Lasso regression with an appropriate value of λ, we select a superset of the true active set S_0^C with high probability. By the sparsity property of the Lasso estimator, it holds that $|\hat{S}^C| \leq \min(n, p)$, hence the name variable screening.

Lemma 7.24. *Assume that*

$$||\hat{\beta}(\lambda) - \beta^0||_1 \leq a, \tag{7.32}$$

with high probability. Then for $C > a$ we have that

$$\hat{S}(\lambda) \supseteq \mathcal{S}_0^C, \tag{7.33}$$

with high probability.

Proof. Let $j \in S_0^C$ then $|\beta_j^0| \geq C$ by definition of S_0^C. Then

$$\begin{aligned}
&||\hat{\beta}(\lambda) - \beta^0||_1 \leq a < C \\
&\implies |\hat{\beta}_j(\lambda) - \beta_j^0| < C \\
&\implies \hat{\beta}_j(\lambda) \neq 0 \\
&\implies j \in \hat{S}(\lambda).
\end{aligned}$$

All equalities and inequalities hold with high probability. □

Variable screening can be useful in numerous applications. Bühlman and Geer (2011) advocate the use of the Lasso estimator with λ sufficiently close to zero and for dimension reduction purposes [25].

7.6.2 Application of Adaptive Lasso to Panel Data

Thus far, we have only considered the (Adaptive) Lasso for regression problems with only a time-dimension. However, panel data[9] is increasingly important in econometric modeling. For the estimation of individual bank-effects in a top-down model for credit loss rates, we consider a panel data model. Since we have time-series of NCO rates for 100 commercial banks in the United States, we might be able to understand for which individual characteristics we need to control in the prediction of rates.

We consider a simple panel data model for n_T observations and n_B individuals of the form,

$$y_{j,t} = \alpha_j + X_t\beta + Z_{j,t}\gamma + \epsilon_{j,t}, \qquad j = 1, \ldots, n_B,\ t = 1, \ldots, n_T, \tag{7.34}$$

where X_t is an $n_T \times p$ matrix of predictors and $Z_{j,t}$ an individual specific $n_T \times q$ matrix of individual effects. The Adaptive Lasso estimator for such a model is given by,

$$\arg\min_{\beta,\gamma} \sum_{j=1}^{n_B}\sum_{i=1}^{n_T}\left(y_i - \sum_{k=1}^{p} X_{i,k}\beta_j - \sum_{j=1}^{q} Z_{i,k,j}\gamma_j\right)^2 + \lambda_1 \sum_{j=1}^{p}\frac{|\beta_j|}{\hat{w}_j^X} + \lambda_2 \sum_{j=1}^{m}\frac{|\gamma_j|}{\hat{w}_j^Z}, \tag{7.35}$$

where λ_1 and λ_2 are tuning parameters for the cross-sectional and time dimensions, respectively. The weight vectors $\hat{w}^X$ and $\hat{w}^Z$ are given by $|\beta^{\text{init}}|^{-\omega}$ and $|\gamma^{\text{init}}|^{-\omega}$, for some $\sqrt{n}$ consistent estimator $(\beta^{\text{init}}, \gamma^{\text{init}})$.

Although a thorough discussion of the application of the Lasso method to different

[9] For a brief introduction to panel data estimation in linear regression, see Appendix B.

types of panel data models is beyond the scope of this thesis, we do include a concise overview of some of the more advanced possibilities for the sake of completeness. Clifford and Souza (2014) propose the use of the Adaptive Lasso method to estimate spatial autoregressive models with exogenous predictor variables[49]. Such a model is of the form,

$$\boldsymbol{y}_t = W^{(1)}\boldsymbol{y}_t + W^{(2)}X_t\beta + \boldsymbol{\epsilon}_t, \quad , t = 1, \ldots, n. \tag{7.36}$$

And the adaptive Lasso minimization problem becomes,

$$\underset{W^{(1)},W^{(2)},\beta}{\arg\min} \sum_{t=1}^{n} ||\boldsymbol{y}_t - W^{(1)}\boldsymbol{y}_t + W^{(2)}X_t\beta||_2^2 + \lambda \sum_{i,j}^{n} \frac{w_{1,ij}}{|\hat{w}_{1,ij}^{\text{init}}|} + \frac{w_{2,ij}}{|\hat{w}_{2,ij}^{\text{init}}|}. \tag{7.37}$$

They derive an oracle inequality and asymptotic consistency properties for this Adaptive Lasso estimator. In their study they apply the estimator to model the dependency between world-wide stock market data, with positive results.

Another effort in applying the adaptive Lasso to panel data with time-varying (or cross-sectional varying) effects is given in [37]. They use well-known Lasso variants such as the *grouped* and *fused* Lasso to model the time-varying behavior of coefficients. Both methods are applicable to feature selection for panel data, but none of the research focuses on the simultaneous estimation of the coefficients of exogenous variables and cross-sectional dependencies. Since the subject of this research project is the assessment of the feasibility of top-down stress testing with Adaptive Lasso, we do not consider time-varying coefficients or correlation structures.

7.7 Summary

We have shown that Adaptive Lasso can automatically select and estimate coefficients from large design matrices with more potential predictors than observations. The selected model is parsimonious and approximately unbiased. Furthermore, the solutions are unique under very flexible conditions. An important property is that variables are selected consistently and the asymptotic distribution of the estimates is as good as if the true model were known in advance. Adaptive Lasso can also be used as a variable screening method that retains substantially relevant predictors with high probability. Lastly, it can be applied to the estimation and selection of panel data models.

Adaptive Lasso is an attractive method for the selection and estimation of top-down models for stress testing. We will use Adaptive Lasso for linear model selection and estimation, as a variable selection method, and in a panel data setting. The models that we construct using Adaptive Lasso are discussed in the following chapter. How we use these models to answer the research questions in Chapter 3, will be discussed in Chapter 9.

Chapter 8

Top-Down Models

The aim of this chapter is to present and explain the method that we use to develop a top-down model for stress testing credit loss rates for individual banks. The purpose of this model is to assess the opportunities of such an exercise regarding prediction accuracy for individual banks, the uses of the inclusion of bank-specific effects, and the aggregation level of the loan portfolio. To that end, we first introduce our modeling approach for the estimation and prediction of NCO rates for averaged loan categories in Section 8.1. We continue with the estimation of NCO rates for fixed loan categories on the individual bank level in Section 8.2, and conclude with methods to aggregate the results for a complete loan portfolio in Section 8.3.

8.1 Averaged NCO Rates by Loan Category

Data-driven model selection consists of obtaining an appropriate initial formulation, and solving the selection, computation, and evaluations problems. In this section we will explain how we implemented the first three steps to learn an appropriate model that can forecast NCO rates conditional on paths of the macro-economy using data-driven methods only. The evaluation of top-down stress testing models will be discussed further in Chapter 9.

The main idea of this research project is to use Adaptive Lasso to generate a hypothesis regarding the model specification of NCO rates, that can be used to predict NCO rates conditional on macro-economic (stress) scenarios, keeping all other circumstances constant. In this section we consider averaged NCO rates, as in equation 3.6, split by loan category. We use Adaptive Lasso both to discover and estimate the model specification for each loan category. Furthermore, we use it as a variable screening method to select substantially relevant predictors, for an auto-regressive model.

The resulting models are used to explore the opportunities and challenges of using top-down stress testing as a complement to bottom-up testing. The predictions of our model allow us to explore the feasibility and limitations of top-down stress testing credit loss rates conditional on paths of the macro-economy. In Section 9.1 we discuss how these models can be used to assess whether Adaptive Lasso can be used to find a parsimonious model, what we can expect of the prediction accuracy of a top-down model, whether auto-regressive effects need to be included, and what the data requirements for top-down stress testing are.

8.1.1 Initial Formulation

In the first step of the data-driven model discovery, the aim is to derive an initial specification of a design matrix X, containing relevant macro-economic variables and transformations thereof. This matrix must be constructed in such a way that all key relevant indicators are included, ensuring that the initial model satisfies the requirements for valid inference that are posed by Doornik and Hendry (2015) [51]. In Section 8.2 we will elaborate on the cross-sectional dimension of the design matrix, which is necessary for the prediction of NCO rates using individual effects, described in Section 3.2.2.

Since we assume that the stress scenario for which we want to predict the NCO rates on loans is given in terms of the variables in the 2014 Federal Reserve stress testing exercise, our initial formulation should at least contain these predictors and transformations thereof. Note that an automated selection method can easily be adapted to include other types of scenarios. It can also be repeated for each quarter, and can adapt to changing circumstances and relations in the macro-economy.

Ideally, we would like to include all significantly relevant transformations of the base predictors in the design matrix. Of course, there are many economically reasonable initial design matrices that we can create. The difficulty in selecting an appropriate initial formulation lies in the fact that we do not want to include too many transformations and interactions of variables to avoid the risk of selecting variables whose relevance cannot be generalized outside the scope of the training data. But we still want to make sure that all relevant variables are included.

Admittedly, the choice for the initial formulation is somewhat arbitrary and further research is needed to establish what kind of initial design matrix is best suited for the top-down stress testing problem, for instance by using the judgment of experts in the field of macro-economics[1]. For our initial formulation, we attempted to specifically take the interpretability of the discovered model into account. A combination of economic, mathematical, and empirical considerations has led to our initial design

[1] An empirical analysis of the impact of the design matrix on the resulting model would require an extra set of data to compare the accuracy of different designs.

matrix.

8.1.1.1 Economic Considerations

Recall the variables of the stress scenario that were introduced in Chapter 5, Commercial Property Price Index (CPPI), Gross Domestic Product (GDP), Dow Jones Industrial Average (DJIA), Disposable Income (DI), House Price Index (HPI), Consumer Price Index (CPI), Chicago Board Options Exchange Market Volatility Index (VIX), Unemployment Rate (UR), and the change in unemployment rate (DUR). We shall from here on refer to these variables as the set of *base predictors*.

We used our economic intuition and published studies to determine potentially relevant transformations of the base predictors. Customarily, the design matrix for the prediction problem in top-down stress testing, is given in terms of independent explanatory variables, which are used to predict the *target* variable, or dependent variable. Lags of the independent and target variables are occasionally considered in time-series regressions. For top-down stress testing most studies include an autoregressive term of the response variable. Research that adopts nonlinear transformations of explanatory variables, such as squares, cubes and interactions are much more scarce. Only Kapinos and Mitnik (2015) [54] use these transformations in the context of top-down stress testing. Nonlinearities around zero are used by Hirtle et al. (2014) [47] by including dummy variables. This allows the model to capture convex relations between the predictor and target variables. In this sense our research presents a novel approach to the top-down stress testing exercise. Although our search for a correct specification is far from exhaustive, we consider far more potential design matrices than any previous top-down stress testing study to our knowledge.

Since the macro-economy may impact NCO rates in a nonlinear manner, we want to consider nonlinear transformations of the base predictors. To capture this behavior we consider lags, indicators, squares, cubes, and interactions between macroeconomic predictors. An intuitive interpretation of the relevance of including these transformations can be as follows.

The response of NCO rates to macro-economic shocks is not necessarily immediate and can last for more than one quarter. The effect can be delayed, or only present when the macro-economic shocks are persistent over time. Hence, we also consider lags of the base predictors, and their average value over the last four quarters. The value or change of a base predictor may indicate the beginning of an economic downturn, or it can be a response. For instance, it is not clear whether a change of the value in a price index is the cause or a consequence of a change in net charge off rates, or perhaps both. Furthermore, the behavior of economic agents is affected by the overall business climate. The response of NCO rates to macro-economic shocks can therefore depend on the value of indicator functions such as $\mathbb{1}_{\{\mathrm{DUR}_t>0\}}$ or $\mathbb{1}_{\{\mathrm{UR}_t>\mathrm{NROU}\}}$[2]. When such indicator functions interact with other base predictors, we may be able to capture nonlinear (or convex) behaviors around zero or the natural rate of unemployment. Moreover, NCO rates may not respond linearly to increasingly large macro-economic shocks. Squares or cubes of macro variables may offer a better description of the relation between these shocks and NCO rates. Also, complicated interactions between different aspects of the macro-economy may be captured by the inclusion of interactions between (transformed) base predictors.

Table 8.1 summarizes the transformations of the base predictors that we considered. We also took interactions (products) between these (transformed) variables into account. This approach is supported by methods that were used in the literature, for instance in Kapinos and Mitnik (2015) and Doornik and Hendry (2015) [54, 51].

Table 8.1: Transformations of Macro-Economic Stress Series

Type of Effect	Transformation
Non-Linearity	$f^{\text{square}}(x_t) = x_t^2$
	$f^{\text{cube}}(x_t) = x_t^3$
Conditioning[1]	$f^{\text{ind.}}(x_t) = \mathbb{1}_{\{x_t>0\}}$
	$f^{\text{neg.ind.}}(x_t) = \mathbb{1}_{\{x_t<0\}}$.
Lag	$f^{\text{lag i}}(x_t) = x_{t-i} \quad i = 1, 2, 3, 4$
Persistence	$f^{\text{average}}(x_t) = \overline{x}_t = \frac{1}{4}\left(x_t + x_{t-1} + x_{t-2} + x_{t-3}\right)$

[1] For the UR_t series we compare with the NROU_t series instead of 0 in the indicator function. For the VIX_t series we use $\mathbb{1}_{\{\mathrm{VIX}>\mathrm{mean}(\mathrm{VIX})\}}$.

8.1.1.2 Mathematical Considerations

Naturally, adding interactions results in a significant augmentation of the size of the design matrix. If we use the Adaptive Lasso method for variable selection, we would like to construct the initial formulation design matrix in such a way that its columns are in general position. Then, by Theorem 7.21 and Definition 7.20, there is a unique solution to the (Adaptive) Lasso minimization problem, even when $p \gg n$. We remark that collinearity may exist in the design matrix. Adaptive Lasso can distinguish between different specifications as long as the columns of X remain in general position.

In order to control for the size of our initial formulation and to preserve uniqueness of the solutions, we want to exclude transformed base predictors if they are in the affine span of other (transformed) base predictors. Testing whether the columns of the initial formulation are in general position is beyond the scope of this thesis, but

[2] Recall that NROU represents the natural rate of unemployment.

we do present some mathematical considerations.

First off, if we include lags $1, \ldots, 4$ for a variable x_t, then $f^{\text{average}}(x_t)$ lies in the affine span of these variables since,

$$f^{\text{average}}(x_t) = \frac{1}{4}(x_t + x_{t-1} + x_{t-2} + x_{t-3}), \tag{8.1}$$

by Definition 7.19. We conclude that, if lags $0, 1, 2$ and 4 of a base predictor are included, than they are in the affine span of the averaged base predictor and the columns of the design matrix are no longer in general position.

Differences such as, $\text{DUR}_t = \text{UR}_t - \text{UR}_{t-1}$ and lags such as, UR_t and UR_{t-1} are typically in general position, since none of the variables lies in the affine span of the other two. This can be seen as follows. Let $f^{\text{lag }0}(x_t)$ and $f^{\text{lag }1}(x)$ be linearly independent $(n_T \times 1)$ vectors with $f^{\text{lag }0}(x) \neq 0$ and $f^{\text{lag }1}(x) \neq 0$. Then there exists no $\alpha \in \mathbb{R}$ such that,

$$\alpha(\pm x_t) - (1 - \alpha)(\pm x_{t-1}) = x_t - x_{t-1}, \tag{8.2}$$

for all $t = 1, \ldots, n_T$. We conclude that DUR_t, UR_t, and UR_{t-1} are typically in general position.

The same is true for positive and negative indicator functions, since there exists no $\alpha \in \mathbb{R}$ such that,

$$\alpha(\pm \mathbb{1}_{\{x_t > 0\}} x_t) + (1 - \alpha)(\pm \mathbb{1}_{\{x_t < 0\}} x_t) = x_t,$$

for all $t = 1, \ldots, n_T$.

We conclude that Adaptive Lasso can be unique when differences of base predictors are included and when positive and negative indicators of base predictors are included. In a sense, this is a remarkable result because it says that Adaptive Lasso can distinguish between different specifications of collinear predictors. Equation 8.1 indicates that care must be taken with the inclusion of both lags and averaged base predictors, since this does not preserve uniqueness.

Another consideration is that the convergence properties of Adaptive Lasso are more favorable for smaller p [14][3]. Therefore, although Adaptive Lasso can handle large initial formulations, it would be prudent to use more specific economic knowledge of the NCO rate process to limit the amount of transformed predictors.

[3] The proof of Theorem 7.13 holds point-wise for fixed p [16]. Extra conditions on the design matrix are required to proof the oracle properties for the $p > n$ case. Derivation of confidence intervals is therefore limited to the $n < p$ case.

8.1.1.3 Design Matrix and Empirical Considerations

The initial formulation that is at the basis of our experiment is presented in Table 8.2. Including the lags, squares, cubes, base interactions, and indicator interactions from this table results in a large design matrix. Since there are 9 base predictors, there are $9 \times 3 = 27$ lagged variables, $9 \times 5 = 45$ averaged lagged variables, $9 \times 8 = 72$ squares, and $9 \times 8 = 72$ cubes. The number of base interactions is equal to $8 \times \sum_{i=1}^{8} i$. For the number of transformed predictors we find that it equals $9 \times 3 \times 3 = 81$, and for the number of averaged transformed predictors we have $9 \times 5 \times 3 = 135$. This brings the total number of indicator interactions to $3 \times (9 \times 2) \times 81 + 3 \times (9 \times 2) \times 135 = 11664$.

Table 8.2: Initial Formulation

Name	Function	
lags	$f_1^{\text{lag}}(x_t) = x_{t-\text{lag}}$	$\text{lag} = 0, 1, 2$
	$f_2^{\text{lag}}(x_t) = \overline{x}_{t-\text{lag}}$	$\text{lag} = 0, 1, 2, 3, 4$
Squares	$f_3^{\text{lag}}(x_t) = x_{t-\text{lag}}^2$	$\text{lag} = 0, 1, 2$
	$f_4^{\text{lag}}(x_t) = \overline{x}_{t-\text{lag}}^2$	$\text{lag} = 0, 1, 2, 3, 4$
Cubes	$f_5^{\text{lag}}(x_t) = x_{t-\text{lag}}^3$	$\text{lag} = 0, 1, 2$
	$f_6^{\text{lag}}(x_t) = \overline{x}_{t-\text{lag}}^3$	$\text{lag} = 0, 1, 2, 3, 4$
Transformed	$g_1(y_t) = f_j(y_t)$	$j = 1, 3, 5$
	$g_2(y_t) = f_j(y_t)$	$j = 2, 4, 6$
Base interactions	$f_i^{\text{lag1}}(x_t) f_j^{\text{lag2}}(y_t)$	$\text{lag1}, \text{lag2} = 0, 1,\ i, j = 1, \ldots, 6,\ x_t \neq y_t$
Indicator interactions[1]	$\mathbb{1}_{\{x_{t-\text{lag}} > 0\}} g_1(y_t)$	$\text{lag} = 0, 1,\ x_t \neq y_t$
	$\mathbb{1}_{\{\overline{x}_{t-\text{lag}} > 0\}} g_2(y_t)$	$\text{lag} = 0, 1,\ x_t \neq y_t$
	$\mathbb{1}_{\{x_{t-\text{lag}} < 0\}} g_1(y_t)$	$\text{lag} = 0, 1,\ x_t \neq y_t$
	$\mathbb{1}_{\{\overline{x}_{t-\text{lag}} < 0\}} g_2(y_t)$	$\text{lag} = 0, 1,\ x_t \neq y_t$
	$\mathbb{1}_{\{x_{t-\text{lag}} > 0\}} x_{t-\text{lag}} g_1(y_t)$	$\text{lag} = 0, 1,\ x_t \neq y_t$
	$\mathbb{1}_{\{\overline{x}_{t-\text{lag}} > 0\}} \overline{x}_{t-\text{lag}} g_2(y_t)$	$\text{lag} = 0, 1,\ x_t \neq y_t$

[1] For the UR_t series we compare with the NROU_t series instead of 0 in the indicator function. For the VIX_t series we compare with the series $\frac{1}{t-1} \sum_{i=1}^{t-1} \text{VIX}_i$ instead of 0 in the indicator function.

The amount of predictor variables in the design matrix is extremely high, and we have that $p \gg n$. The idea of (Adaptive) Lasso, is that it will only select a parsimonious model consisting of the most relevant variables. To empirically test our initial formulation we generated samples for simple models present in the initial formulation and added normally distributed noise. Furthermore, we tested the initial formulation for sample paths of an auto-regressive model. That is, we generated sample paths of the following simple models,

$$y_t = \mathrm{DUR}_t + \epsilon_t, \tag{8.3}$$

$$y_t = \epsilon_t, \tag{8.4}$$

$$y_t = y_{t-1} + \mathrm{DUR}_t + \epsilon_t, \tag{8.5}$$

where $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$. We assessed which variables were selected by Adaptive Lasso, based on the generated data. The results of this small experiment are displayed in Table 8.3.

Table 8.3: Initial Formulation Test Results

model	nr. of samples	error size[1]	selected variables	frequency[2]
Equation 8.3	65	10%	DUR_t	10/10
Equation 8.3	65	25%	DUR_t	3/10
Equation 8.3	96	25%	DUR_t	8/10
Equation 8.4	65		none	10/10
Equation 8.5	65	10%	$\overline{\mathrm{DUR}}_{t-\mathrm{lag}}$	5/10
			interactions with $\overline{\mathrm{DUR}}_t$	6/10
			other	0/10

[1] Percentage of the range of the data without error.
[2] Number of times the variable was selected, compared to the total number of tests.

In the first experiment, The Adaptive Lasso method was able to retrieve the correct variables from the design matrix, provided that the noise was kept relatively low compared to the strength of the signal. When the noise is larger, more samples are needed for reliable variable selection. For the second experiment, no variables were selected. And in the final experiment, where the true model was not present in the data, Adaptive Lasso selected a variety of interactions with lags of averaged values of DUR_t and interactions of these averaged values with other predictors. These experiments suggest that the initial design as described in Table 8.2 can select the correct model with relatively high probability for data generated from simple specifications. A second empirical consideration is that we repeated part of our analysis for a much simpler initial formulation, including only lags, fewer indicator functions, squares, and a limited amount of interactions. This resulted in a design matrix with ~ 500 predictor variables. The resulting models were similar in the sense that their economic interpretation was comparable to that of the full initial design, and the typical number of selected variables was low (≈ 5), but the predictions were slightly less accurate than those for the model based on the full design[4].

[4]For future research it would be interesting to thoroughly assess the limitations and opportunities of the specification of the initial formulation.

8.1.2 Selection Problem

For the selection problem there are two main considerations. In the first place, relevant macro-economic predictors and transformations that influence average NCO rates should be selected from the set of candidate predictor variables in the initial formulation. And secondly, we want to determine whether we should assume a linear or auto-regressive model structure for the top-down modeling of NCO rates on loan categories. In this section we consider averaged NCO rates on loan categories $P_i \in \mathcal{P}$ but we drop the subscript i from our notation.

Most papers about top-down stress testing, consider an auto-regressive model with exogenous predictors, which we shall refer to as an ARX-model. Such a model is of the form,

$$\mathrm{NCO}_t = \alpha + X_t\beta + \sum_{i=1}^{p} \phi_i \mathrm{NCO}_{t-i} + \epsilon_t, \tag{8.6}$$

with $p \in \mathbb{N}$ and ϵ_t are i.i.d random variables with mean zero and variance σ^2. For a proper introduction to time-series modeling, we refer to Appendix A. Although the NCO rate processes for different loan categories in Figure 6.3 indeed seem to be of an auto-regressive nature, perhaps this behavior can also be explained by a dependency on the macro-economy, which also behaves auto-regressively. This last option is more favorable, since it can give more accurate predictions when we assume that the future state of the macro-economy is known, which is the case in a stress testing context.

During the research project, we considered two model types for the dependency of NCO rates on macro-economic variables, a linear and an auto-regressive model. For the linear model we used Adaptive Lasso to select variables and estimate coefficients in one step. For a model with auto-regressive effects we used Adaptive Lasso as a variable screening method to select substantially relevant predictors. In a second step we estimated the coefficients of an auto-regressive model. The aim of this second approach is to determine whether an auto-regressive model is more appropriate for top-down models of credit loss rates than a linear model. Because one of the aims of this research project is to investigate the possibility of predicting rates for stress horizons far beyond 9 quarters, we take a closer look at linear models.

8.1.2.1 Linear Model

Our main model for NCO rates is of the form,

$$\mathrm{NCO}_t^{\mathrm{av}} = \alpha + X_t\beta + \epsilon_t, \tag{8.7}$$

where ϵ_t is a white noise process with mean zero and constant variance σ^2 and X is the design matrix described by the initial formulation in Section 8.1.1. The linear model assumes that the deterministic part of the NCO rates is completely determined by the predictor variables in the design matrix, and the remaining random part $\epsilon = (\epsilon_1, \ldots, \epsilon_{n_T})$ are i.i.d. random variables with mean zero. We use the Adaptive Lasso method to obtain a sparse estimate $\hat{\beta}^{\text{alasso}}$. We use the Lasso estimator to determine the weight vector $\hat{w} = |\hat{\beta}^{\text{init}}|^{-\gamma}$ of Adaptive Lasso.

Adaptive Lasso shrinks coefficients that are less relevant to zero, thus reducing model complexity and yielding a sparse model. In the previous chapter, we have shown that the solution is sparse, continuous and approximately unbiased. We also proved that the estimates are unique when the column of the design matrix are in general position by Lemma 7.21. Moreover the adaptive Lasso procedure possesses the oracle properties by Theorem 7.13. The method can adequately select the true model parameters from a large set of potentially relevant predictor variables, and gives asymptotically unbiased estimates at the same time. A major advantage of this model, is that it can be easily adapted for different kinds of stress scenarios and changes in the market, and it is completely automated. Moreover, it gives a parsimonious description of the movements of NCO rates conditional on exogenous paths of the macro-economy. These models can be verified and interpreted from a theoretical perspective.

8.1.2.2 Autoregressive Moving Average Model

There is a large amount of possible predictor variables in the initial formulation, from which we want to select a parsimonious and relevant set that gives accurate predictions. Here, we use Adaptive Lasso as a variable screening method. By Lemma 7.24 substantially relevant predictors are selected with high probability. We shall refer to this set by X^{screen}. Subsequently, we estimate an auto-regressive model with exogenous predictors X^{screen}.

Recall that most models that we encountered in the literature use an auto-regressive model of the form,

$$\text{NCO}_t = \alpha + \beta X_t^{\text{screen}} + \text{NCO}_{t-1} + \epsilon_t, \tag{8.8}$$

where ϵ_t are i.i.d. random variables with mean zero and variance σ^2. The term NCO_{t-1} is an auto-regressive term, taking into account the previous values of the NCO rate[5]. In practice it will be difficult to determine the amount of autocorrelation that is due to a dependency on autocorrelated macro-economic variables, and

[5]See Appendix A for an introduction to auto-regressive models and overall background information of time-series modeling.

autocorrelation of the NCO_t process itself. Another major drawback of this model is that $\hat{\beta}$ can only be interpreted conditional on the previous state of the NCO rates.

Our approach is to model the error terms as an auto-regressive moving average process directly. We refer to such a model as a RegARMA(p, q) model. It can be written as follows,

$$\begin{aligned} Y_t &= \alpha + \beta X_t^{\text{screen}} + \eta_t \\ \eta_t &= \sum_{i=1}^{p} \phi_i \eta_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t, \end{aligned} \tag{8.9}$$

where ϵ_t are i.i.d. random variables with mean zero and variance σ^2. The parameters ϕ_i and θ_i represent the coefficients of the auto-regressive lags and the moving average terms, respectively. The number of autoregressive lags p and the number of moving average terms q is chosen by selecting the model with the lowest value of the Bayesian Information Criterion (BIC),

$$\text{BIC} = k \cdot \log(n) - 2\log(L), \tag{8.10}$$

where k is the number of estimated parameters in the model and L is the likelihood function[6].

A linear model with ARMA errors can be selected and estimated in one step by an adjusted version of Adaptive Lasso [33]. Our approach estimates the model in two steps. First the exogenous predictors are selected by Adaptive Lasso variable screening, and then the coefficients of the RegARMA(p,q) are estimated. Because the adjusted version of Adaptive Lasso is not computationally efficient, we decided to use the second approach for our research project.

8.1.3 Computational Problem

Convex minimization problems can be solved efficiently with coordinate descent algorithms[25]. Conveniently, the lasso minimization problem in definition 7.3 is a convex problem. In this section we discuss how the Adaptive Lasso estimator can be computed.

8.1.3.1 Data Centering

For Adaptive Lasso methods, we assume that the data in the design matrix is centered and scaled. All predictors are standardized before estimation by computing

[6]For background on model selection criteria such as BIC and likelihood functions we refer to Appendix B.

for each column $X^{(j)}$, $j = 1, \dots, p$,

$$X^{(j)}_{\text{scaled}} = \frac{X^{(j)} - \frac{1}{n}\sum_{i=1}^{n} X_{i,j}}{\text{sd}(X^{(j)})}, \tag{8.11}$$

where $\text{sd}(X^{(j)})$ denotes the standard deviation of $X^{(j)}$. Scaling is necessary to ensure that the relative size of the predictor variables does not influence the penalty function.

8.1.3.2 Cyclical Coordinate Descent

Solutions to convex minimization problems can be obtained with cyclical coordinate descent algorithms. The idea is to minimize a convex objective function along each coordinate iteratively. To obtain the Lasso estimator, we want to find $\beta \in \mathbb{R}^p$ that minimizes equation 7.4 and satisfies the constraint. We use a cyclical coordinate descent algorithm that minimizes the convex unconstrained problem along dimensions $j = 1, \dots, p$ iteratively. Lemma 8.1 shows that the cyclical coordinate descent algorithm converges to the global minimum of the Lasso objective function. Below we describe the algorithm in more detail.

Assume that $\lambda > 0$ is fixed and write,

$$f(\beta) = \frac{1}{2n}||Y - X\beta||_2^2 + \lambda||\beta||_1. \tag{8.12}$$

Note that $f(\beta)$ is a convex function by Lemma 7.15. We start with an initial estimator $\beta^{(0)}$, and calculate for $k = 1, 2, \dots$,

$$\begin{aligned}
\beta_1^{(k)} &= \arg\min_{\beta_1} f(\beta_1, \beta_2^{(k-1)}, \dots, \beta_p^{(k-1)}) \\
\beta_2^{(k)} &= \arg\min_{\beta_2} f(\beta_1^{(k)}, \beta_2, \dots, \beta_p^{(k-1)}) \\
&\vdots \\
\beta_p^{(k)} &= \arg\min_{\beta_p} f(\beta_1^{(k)}, \beta_2^{(k)}, \dots, \beta_p).
\end{aligned}$$

That, is we minimize the objective function along each dimension. The process is repeated until $||\beta^{(k)} - \beta^{(k-1)}|| < \epsilon$, for some ϵ sufficiently small. The minimum in one coordinate direction is found by a line search. When the coordinate directions are computed cyclically throughout the procedure, we speak of a cyclical coordinate descent algorithm.

The convergence of coordinate descent algorithms for differentiable convex optimization problems are well established[11]. The Lasso cost function however, is not differentiable at zero. Lemma 8.1 below states that coordinate descent also converges for the Lasso minimization problem.

Lemma 8.1. *(Coordinate Descent for Lasso) The value of β for which,*

$$\frac{1}{2n}||Y - X\beta||_2^2 + \lambda||\beta||_1, \tag{8.13}$$

obtains it global minimum can be computed by coordinate descent.

Proof. For the proof of this result we refer to [20]. □

This shows that we can solve the Lasso minimization problem for fixed $\lambda > 0$. Thus, we have an estimator $\hat{\beta}(\lambda)$ as a function of λ. Below we describe how a suitable λ can be chosen along a so-called *regularization* path.

8.1.3.3 Cross-Validating the Tuning Parameter

For the computation of Lasso estimates we use the R-package glmnet [44][7]. At the moment of writing, this is one of the fastest algorithms available to compute Lasso estimates. The package uses a cyclical coordinate descent algorithm along a regularization path of the tuning parameter λ.

In the literature on Lasso algorithms, there exist several suggested methods to find the optimal λ. In this section we describe the method that is used in the glmnet package. This algorithm calculates the solutions to the Lasso minimization along a regularization path for λ, which is determined by first selecting a tuning parameter $\lambda^{\max}$ sufficiently large, such that $\hat{\beta} = 0$. The minimal tuning parameter of the regularization path is then calculated by $\lambda^{\min} = \epsilon\lambda^{\max}$. The sequence of the regularization path is constructed by taking K values of $\lambda \in [\lambda^{\min}, \lambda^{\max}]$ on the logarithmic scale [44]. In our case, $\epsilon = 0.001$ and $K = 100$.

Via the cyclical coordinate descent algorithm, the minimization problem is solved for each λ on the regularization path described above. The optimal value for the tuning parameter is determined by 10-fold *cross-validation*.

For each λ on the regularization path, we can determine the so-called cross-validation error of the Lasso estimate. We calculate this error by separating the estimation data into ten samples (or folds) of equal size. That is, we separate n_T observations in ten equal sets, and let $Y(k)$ and $X(k)$ denote the observations in the kth sample. We retain the sample k for validation. We then use the nine remaining samples to

[7] The Statistics and Machine Learning Toolbox in Matlab also has an excellent algorithm available, which has an option to use parallel computing.

compute the Lasso estimator $\hat{\beta}_{(-k)}$. Where the subscript $(-k)$ indicates that the estimator is calculated on all training data accept the kth sample that was retained for validation.

For all λ on the regularization path, the mean square prediction error is computed on the kth validation sample,

$$\text{MSE}_k(\lambda) = ||Y(k) - X(k)\hat{\beta_{(-k)}}||_2^2. \tag{8.14}$$

This procedure is then repeated for each sample $k = 1, \dots, 10$, which allows us to compute the cross-validation error,

$$\text{CV}(\lambda) = \sum_{k=1}^{10} \text{MSE}_k(\lambda). \tag{8.15}$$

The optimal tuning parameter λ^* is that value of λ for which the cross-validation error is minimal,

$$\lambda^* = \arg\min_{\lambda} \text{CV}(\lambda). \tag{8.16}$$

The Lasso solution can now be obtained by solving,

$$\hat{\beta}^{\text{lasso}} = \arg\min_{\beta} \frac{1}{2n}||Y - X\beta||_2^2 + \lambda^*||\beta||_1. \tag{8.17}$$

8.1.3.4 Computation of the Adaptive Lasso

As discussed in the previous paragraph, there are several fast packages, based on coordinate descent algorithms, available that solve the Lasso minimization problem. However, we would like to minimize *Adaptive* Lasso problems. By Proposition 8.2 we can use existing procedures for standard Lasso estimation to obtain Adaptive Lasso estimators.

Proposition 8.2. *Let $\hat{w}_j$ be the weights used for Adaptive Lasso. We define $\beta_j^* = \beta_j \hat{w}_j$ and rescale $X_{ij}^* = X_{i,j}/\hat{w}_j$ for all $j = 1, \dots, p$. Then minimizing,*

$$\sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j X_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \hat{w}_j, \tag{8.18}$$

is equivalent to minimizing,

$$\sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j^* X_{ij}^* \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j^*|, \tag{8.19}$$

Proof. Since $\hat{w}_j$ is non-negative, plugging $\beta_j^* = \beta_j \hat{w}_j$ and $X_{ij}^* = X_{ij}/\hat{w}_j$ into equation 8.18 yields equation 8.19. □

Remarkably, Adaptive Lasso estimators can be obtained by an adaptive rescaling of the predictor variables. The Adaptive Lasso estimator $\hat{\beta}^{\text{alasso}}$ can be computed by rescaling X by $\hat{w} = |\hat{\beta}^{\text{lasso}}|^{-\gamma}$ as in Proposition 8.2. Then the Lasso algorithm can be used on the scaled design matrix to obtain a solution $\hat{\beta}^{\text{lasso}*}$. It follows that $\hat{\beta}^{\text{alasso}} = \hat{\beta}^{\text{lasso}*}/\hat{w}$. An algorithm that can be used to compute Adaptive Lasso estimators for λ and γ with minimal cross-validation errors can be found in Algorithm 2 in Appendix C.

In this case, the optimal value for the tuning parameter λ can be obtained by determining which λ on the regularization path gives the lowest cross-validation error for Adaptive Lasso predictions. Moreover, the optimal value for γ can also be determined by cross-validation. An algorithm that computes the Adaptive Lasso estimator for $(\lambda^{\text{opt}}, \gamma^{\text{opt}})$, is explained in pseudo-code in Algorithm 2 in Section C.3.

8.2 NCO Rates with Individual Effects

In the previous section we described and motivated the method that we use to estimate the impact of macro-economic developments on NCO rates of averaged loan categories. In this section we consider a panel data model with bank-specific effects, as explained in Section 3.2.2. For the prediction of Net Charge Offs rates for individual banks, we would like to capture the variability in the cross-sectional dimension. To that end, we first want to assess whether bank-specific variables help explain NCO rates on loan categories for individual banks. Furthermore, we want to determine whether scaling factors for individual banks can help explain variability between NCO rates on loan categories for individual banks.

8.2.1 Initial Formulation

For the macro-economic dimension of individual bank NCO rate estimation, we consider the initial formulation of Section 8.1.1, and refer to it as M_t.

For the cross-sectional dimension of the design matrix we consider the CAMELS criteria in Table 5.2, which are used to assess the condition of financial institutions. Recall that the CAMELS criteria are given by the following eight financial measures, capital ratio, provision rate, non-interest profit, return on equity, net interest margin, total liquid assets, total loans, and loan growth. Recall that since NCO rates and the provision rate are linked directly through some provisioning scheme [47], we choose to discard provision rate as an explanatory variable. From here on, we denote the time-series of these criteria at time t, for banks $b_j : j = 1, \ldots, n_B$ by $B_{j,t}$. Assume that we have NCO rates and CAMELS data for n_B banks over n_T quarters, then B is an $(n_B \times n_T \times 8)$ matrix, since there are 8 bank-specific variables in the CAMELS criteria.

An important goal of the individual estimation exercise is to determine whether bank-specific variables should be included in a top-down stress testing model for credit losses on loan categories. A model that better describes the behavior of individual banks' NCO rates, could make better predictions conditional on macro-economic paths keeping other circumstance constant, because it allows us to control for some of these remaining circumstances. Here, we do not include (non-linear) transformations and interactions with bank-specific variables. A carefully chosen specification of the bank-specific dimension of the top-down stress testing estimation problem could be an interesting topic for future research.

8.2.2 Selection Problem

Because we have three-dimensional data, which includes both bank-specific variables for a large number of banks and transformed macro-economic variables, a panel data model could be an appropriate choice. In such a model the estimation of the coefficients of macro-economic predictors and those of bank-specific predictors is performed simultaneously. With the inclusion of bank-specific CAMELS criteria $B_{j,t}$, we assume that the NCO rates of banks $b_j : j = 1, \ldots, n_B$ satisfy the following panel data model specification,

$$\mathrm{NCO}_{j,t} = \alpha_{j,t} + c_{j,t} \left(M_t \beta + B_{j,t} \gamma + \epsilon_t \right), \tag{8.20}$$

where the errors ϵ_t are i.i.d. random variables with mean zero. Note that the above model is for loan category $P_i \in \mathcal{P}$, where we drop the subscript i from the notation for $\mathrm{NCO}_{i,j,t}$, $\alpha_{i,j,t}$, and $c_{i,j,t}$. We refer to $\alpha_{j,t}$ and $c_{j,t}$ as individual scaling factors.

To simplify the estimation problem, we assume that the time-series for the different bank NCO rate processes are not correlated, except for their common dependency on macro-economic variables. Moreover, we assume that the rates of different loan categories for the same bank are not correlated, except for their common dependency on the macro-economy and bank-specific variables. Another important assumption is that the coefficients of the (transformed) variables are time-independent. These assumptions are implicit in all top-down models that we encountered in our literature study. Under these assumptions, the model specification is given by equation

$$\mathrm{NCO}_{j,t} = \alpha_j + c_j \left(M_t \beta + B_{j,t} \gamma + \epsilon_t \right), \tag{8.21}$$

The future values of the bank-specific variables are, unlike the macro-economic predictors, not assumed to be known at the time of prediction. Therefore it might be prudent to apply a lag of length h to B_t. However, this reduces the length of the time-series available for estimation significantly. This has a particular adverse effect on the estimation, since the data at the beginning of the time-series contains important information about the developments in the economically stressed period between 1991 Q1 and 1993 Q4, as can be observed in Figure 6.3. Hence, we do not adjust the data for the lag length, but take the CAMELS criteria as a constant for our predictions. In other words, when we use the estimated model $(\hat{\alpha}_j, \hat{c}_j, \hat{\beta}, \hat{\gamma})$ for prediction, we freeze the values of the bank-specific predictors. That is, we calculate,

$$\mathrm{NCO}_{j,t+h} = \hat{\alpha}_j + \hat{c}_j \left(M_{t+h}\hat{\beta} + \hat{\gamma} B_t \right) + \epsilon_{j,t+h}, \tag{8.22}$$

where the errors $\epsilon_{j,t}$ are i.i.d. random variables with mean zero.

To select and estimate the model for an arbitrary loan category, we minimize the following Adaptive Lasso problem,

$$\sum_{t=1}^{n_T} \sum_{j=1}^{n_B} \left(\mathrm{NCO}_{j,t} - \alpha_j - c_j X_{j,t}\beta_j - c_j B_{j,t}\gamma\right)^2 + \lambda_1 ||w_M \beta||_1 + \lambda_2 ||w_B \gamma||_1, \tag{8.23}$$

where the weight vectors w^M and w^B are given by

$$\hat{w}^M = |\beta^{\mathrm{lasso}}|^{-\omega}, \qquad \hat{w}^B = |\gamma^{\mathrm{lasso}}|^{-\omega}, \tag{8.24}$$

where ω is a parameter controlling the weights.

Alternatively, excluding bank-specific predictors, we only consider individual scaling factors in the model,

$$\mathrm{NCO}_{j,t} = \alpha_j + c_j \left(M_t \beta + \epsilon_t \right). \tag{8.25}$$

In this case, we obtain the estimator $\hat{\beta}$ by using the model for averaged loans in Section 8.1.

8.2.3 Computational Problem

We describe how the computational methods in Section 8.1.3 can be used for the estimation of the models in this section. First, we estimate α_j and c_j for banks $b_j : j = 1, \ldots, n_B$ by normalizing and rescaling the NCO data for the time-series of each bank in the test set,

$$\begin{aligned}
\hat{\alpha}_j &= \text{mean}(\text{NCO}_j) \\
\hat{c}_j &= \text{sd}(\text{NCO}_j) \\
\overline{\text{NCO}}_j &= \frac{\text{NCO}_j - \alpha_j}{c_j}.
\end{aligned} \tag{8.26}$$

The panel data model with bank-specific effects in equation 8.22 can then be estimated by applying the Adaptive Lasso method in the previous section to the matrix,

$$X = \begin{pmatrix}
M_{1,1} & \cdots & M_{1,p} & B_{1,1,1} & \cdots & B_{1,1,9} \\
M_{2,1} & \cdots & M_{2,p} & B_{1,2,1} & \cdots & B_{1,2,9} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
M_{96,1} & \cdots & M_{96,p} & B_{1,96,1} & \cdots & B_{1,96,9} \\
M_{1,1} & \cdots & M_{1,p} & B_{2,1,1} & \cdots & B_{2,1,9} \\
M_{2,1} & \cdots & M_{2,p} & B_{2,2,1} & \cdots & B_{2,2,9} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
M_{96,1} & \cdots & M_{96,p} & B_{2,96,1} & \cdots & B_{2,96,9} \\
\vdots & & \vdots & \vdots & & \vdots \\
M_{96,1} & \cdots & M_{96,p} & B_{100,96,1} & \cdots & B_{100,96,9}
\end{pmatrix}, \tag{8.27}$$

with response variable

$$y = (\text{NCO}_{1,1}, \ldots \text{NCO}_{1,96}, \text{NCO}_{2,1}, \ldots, \text{NCO}_{2,96}, \ldots \text{NCO}_{100,96})^\top. \tag{8.28}$$

The tuning parameters λ_1 and λ_2 can be estimated by using cross-validation. For instance, this can be done by re-weighting the coefficients for the CAMELS criteria according to the scheme described in Proposition 8.2. The cross-validation technique can then be used to determine the optimal value for λ_1 and λ_2. Note that this minimization problem has two tuning parameters, and both need to be optimized using cross-validation. Although the problem is still convex and can be solved efficiently, the cross-validation technique to determine the optimal value of tuning parameters is rather slow. For simplicity and a reduction in computation time, we solved the minimization problem for $\lambda_1 = \lambda_2$ only.

We need to pay attention to the cross-validation procedure. Recall that the cross validation technique to determine the optimal λ, splits the sample into $k = 10$ equal parts. The model is estimated on $k-1$ parts and the mean square error is determined on the kth part. The tuning parameter is chosen such that the mean square error is smallest on the k folds.

For panel data, the creation of folds becomes more complicated than for a linear model. If we construct folds in the cross-sectional dimension, the tuning parameter is chosen so that the cross-sectional mean square error of prediction is optimal. If we construct the folds in the time-dimension the focus of the cross-validation is on future predictions. To estimate whether the CAMELS criteria help explain cross-sectional variability, we create folds for cross-validation by splitting the time-series for banks into k folds. For prediction of NCO rates given macro-economic developments, we create folds for cross-validation by splitting the data into k folds in the time-dimension.

8.3 NCO Rates for Individual Banks

In the previous section we described how bank-specific estimates for NCO rates on specific loan categories can be obtained with and without individual effects. Here we describe how the predictions for different loan categories can be aggregated to determine the NCO rate on the total loan portfolio of an individual bank.

Recall the aggregation levels that we introduced in Table 6.2. In Section 8.2.2 we showed two ways to obtain forecasts of NCO rates for loan categories at aggregation levels A_1, A_2, and A_3. Individual scaling factors can be obtained by equation 8.26. These can be used to scale the predictions for averaged loan categories, which are obtained by estimating $\hat{\beta}^{\text{alasso}}$ for the models in equations 8.7 and 8.9. These estimators are then used in equation 8.25. The second approach involves obtaining forecasts for the estimated panel data model in equation 8.22.

When we have obtained predictions for all loan categories at one of the aggregation levels, we want to combine these to give a forecast for the NCO rate for the complete loan portfolio of an individual. A simple method to aggregate NCO predictions for different categories is to add the predicted values as follows. Let a be the number of categories at the aggregation level of the predictions, and $s \in \{1, \dots, h\}$, where h is the stress horizon. Then the predicted total NCO rate is a weighted average of the predicted rate per loan category and banks $b_j : j = 1, \dots, n_B$,

$$\widehat{\text{NCO}}^{\text{tot}}_{j,t+s} = \sum_{i=1}^{a} \widehat{\text{NCO}}_{i,j,t+s} \frac{\text{TL}_{i,j,t}}{\text{TL}^{\text{tot}}_{j,t}}. \tag{8.29}$$

Because we assume that the loan composition of the individual remains the same over the stress horizon, we use the ratios of the loan categories at the time that the forecast is made.

Chapter 9

Evaluation of the Method

The main purpose of the method that we developed is to establish the feasibility of using a top-down model for stress testing. A second aspect is to determine what can be expected of such a model in terms of prediction accuracy. Finally, if the method is to be used to perform stress tests, the validity and explanatory power of the model is to be tested.

For the evaluation of a top-down stress testing model, it is important to understand the strengths and weaknesses of the model, and ultimately, for their practical use, the predictions of the model need to be equipped with a measure of uncertainty. These measures depend on the application of the stress test results. In particular, the intention is to assess the opportunities, challenges and limitations for top-down stress testing for Rabobank. Note that this means that the model on data from the United States might not be representative of the relevant features of a large Dutch bank such as the Rabobank.

An assessment of the quality of the model is crucial for automated feature selection procedures in particular, since these models are prone to increase model risk. In the automated model selection scheme of Section 7.1, the evaluation problem is the final step in completing the procedure for automated feature selection.

When a dataset is used to find a suitable hypothesis or model, we can not use the same data to test the validity, error margins or statistical significance of such a hypothesis. Therefore we use measures that assess the prediction accuracy on true out-of-sample test data. The data in the test set cannot be used for any part of the estimation procedure. To that end, the NCO rate data from 1991 Q1 to 2014 Q4, 96 quarters in total, is divided in a *training* and a *test* set for each loan category and all banks, where the financial crisis of 2007-2009 is included in the test set. This allows us to assess the prediction accuracy of the model, in a period where both a recession and a boom take place. The training set consists of the first 65 quarters (1991 Q1 to 2007 Q1) and the test set contains the last 31 quarters (2007 Q2 to

2014 Q4) of NCO rate data.

9.1 Average Predictions

For the evaluation of predictions on averaged loan categories, we take several criteria into account. The first is economic significance of the selected variables, the second is prediction accuracy of the estimated model, and lastly we check whether the model explains the internal variation in the data and satisfies the assumptions of the model. We compare the results for the Adaptive Lasso and the RegARMA model, and we use a simple model to benchmark the results.

9.1.1 Benchmarks

We introduce a benchmark model so that the results of both the linear Adaptive Lasso and the RegARMA model can be compared to a basic model, which is comparable to current top-down modeling practices. Our benchmark model is based on the model that is used in Hirtle et al.(2014)[47]. Unfortunately, we cannot directly use their model specification as a benchmark, because the model features were selected based on the same data that we used from 1991 Q1 to 2013 Q3, and therefore we cannot compare its out-of-sample performance to the results of our model.

To circumvent this, we replicate part of their approach to the modeling of NCO rates. We construct a much smaller and simpler design matrix than the initial formulation of Section 8.1.1, and similar to the variables that were considered in [47]. We start with the macro-economic variables that are specified in the Federal Reserve's stress testing exercise, and focus on the set of base predictors. We include the following indicator functions to capture nonlinear behavior around zero or convexity,

$$
\begin{aligned}
&\mathbb{1}_{\{\mathrm{UR}_t > \mathrm{NROU}_t\}}\mathrm{UR}_t, && \mathbb{1}_{\{\mathrm{GDP}_t < 0\}}\mathrm{GDP}_t,\\
&\mathbb{1}_{\{\mathrm{HPI}_t < 0\}}\mathrm{HPI}_t, && \mathbb{1}_{\{\mathrm{CPPI}_t < 0\}}\mathrm{CPPI}_t,\\
&\mathbb{1}_{\{\mathrm{INF}_t < 0\}}\mathrm{INF}_t, && \mathbb{1}_{\{\mathrm{DJI}_t < 0\}}\mathrm{DJI}_t,\\
&\mathbb{1}_{\{\mathrm{DUR}_t > 0\}}\mathrm{DUR}_t, && \mathbb{1}_{\{\mathrm{VIX}_t > \mathrm{mean(VIX)}\}}\mathrm{VIX}_t,\\
&\mathbb{1}_{\{\mathrm{DI}_t < 0\}}\mathrm{DI}_t.
\end{aligned}
$$

These 18 variables make up the benchmark design matrix, which we shall refer to as X^{small}.

The goal is to include those variables from the predictor set that are statistically significant, and result in the lowest value for the Bayesian Information Criterion (BIC), resulting in a predictor matrix $X^{\min(\text{BIC})}$. The structural form of the model is taken to be linear[1],

$$\mathrm{NCO}_{i,t} = \alpha_i + \beta X_t^{\min(\mathrm{BIC})} + \epsilon_{i,t}, \tag{9.1}$$

where $\epsilon_{i,t}$ are white noise.

We use a computationally-intensive greedy backward elimination algorithm, which is displayed in Algorithm 1, to find the model with the lowest BIC value for each loan category, and select it as our benchmark model. This means that we start with a model which includes all the macro-economic predictor variables in X^{small}. Then we remove that predictor variable whose removal leads to the greatest decline of the BIC score if it improves (lowers) the score with respect to the model with one extra variable included. This process is iterated until removal of another variables leads to an increase in the BIC score. The resulting model is of the form of equation 9.1 and serves as a benchmark for the linear Adaptive Lasso and RegARMA models.

Data: design matrix X of size $n \times p$, response variable y of size $n \times 1$
Result: benchmark model β

initialize;
estimate linear/arma/regarma model using y and X;
bic.init $\longleftarrow$ bic value of the model;
bic $\longleftarrow$ bic.init;
removed $\longleftarrow$ []

while *min(bic)* $<$ *bic.init* **do**
    $X \longleftarrow X$ without column *removed*;
    bic $\longleftarrow$ zeros(size$(X, 2)$);
    foreach *column i of X* **do**
        $\beta \longleftarrow$ model for y and X without column i;
        bic(i) $\longleftarrow$ bic value of β;
    end
    removed $\longleftarrow$ index of min(bic);
end
$\beta \longleftarrow$ estimated model for y and X

Algorithm 1: Backward Elimination

It is convenient to compare a model's performance against a *lagging* model, which gives the current NCO rate as a prediction for the stress horizon. The lagging model

[1]The linear model serves as a benchmark to Adaptive Lasso, which is estimated on a much larger design matrix. As a benchmark to the literature, we also considered an auto-regressive model specification as in Hirtle et al. [47]. Empirically, we observed that the prediction accuracy for an auto-regressive structure was generally comparable to that of the linear form.

is defined as,

$$\mathrm{NCO}_{t+h} = \mathrm{NCO}_t. \tag{9.2}$$

Finally, we compare the predictions of the adaptive Lasso, RegARMA, and benchmark model with the predictions of a *null* model, which is given by,

$$\mathrm{NCO}_{t+h} = \frac{1}{t}\sum_{j=1}^{t} \mathrm{NCO}_j. \tag{9.3}$$

9.1.2 Selected Variables

We use a data-driven approach to select a model for the prediction of NCO rates on specific loan categories. An important part of the evaluation of the discovered model is to verify that the selected variables are economically significant, and that the signs of the coefficients are interpretable from an economical perspective.

In a completely data-driven modeling process there is the danger of spurious regression. That is, we are at risk of including variables into our model that correlate with the target (or outcome) variable within the training set, but are not related in general. If this is the case, then the prediction accuracy of the method is greatly reduced because the model cannot be generalized beyond the scope of the training set. Therefore a hybrid approach of automated selected features and a check for economical significance based on expert opinion is a prudent course of action.

9.1.3 Prediction Accuracy

The prediction accuracy of the model is determined by estimating the model on the training set and determining its prediction Mean Absolute Error (MAE) and Mean Square Error (MSE) on the test set. Let f denote a model that predicts response variables y_i from predictors x_i. In our case f represents either the linear Adaptive Lasso, RegARMA, or benchmark model. The prediction error statistics are calculated as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |f(x_i) - y_i|, \tag{9.4}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n} (f(x_i) - y_i)^2. \tag{9.5}$$

Models whose predictions are close to the observed value and thus have low values for MSE and MAE, are generally preferred, because of their superior predictive performance. It can be argued however, that for a model which is intended for stress testing, where severe but plausible scenarios are considered, is mainly intended as a warning system and not used for exact predictions.

When we want to compare the prediction accuracy for different individual banks, the NCO rates can be on a different scale. In such cases, we compute the normalized mean absolute prediction error and normalized square prediction eror to compare the prediction errors for different institutions. These measures are given by,

$$\text{NMAE} = \frac{\text{MAE}}{\max(\text{NCO}) - \min(\text{NCO})}, \tag{9.6}$$

$$\text{NMSE} = \frac{\text{MSE}}{\max(\text{NCO}) - \min(\text{NCO})}. \tag{9.7}$$

Evaluating the prediction accuracy on the test set resembles the application and purpose of stress testing most closely, but it leaves us with only one test sample for the procedure. Since we are particularly interested in the prediction accuracy for a forecast horizon $h = 9$, we also test prediction accuracy by constructing 9-quarters-ahead forecasts. A typical technique is the rolling window forecasting method, where the model is estimated on n data points. The 9-quarters-ahead forecast for $t = t_i$ is obtained by estimating the model on the time points $t = (t_i - h - n), \ldots, (t_i - h)$, and then evaluating the resulting model on t_i.

Figure 6.3 indicates that, in our case, much information is lost when we do not use the first data-points that are available. Therefore we use, what we call, an *extending* estimation window. This means that the forecast for $h = 9$ at time t_i is obtained by estimating the model on $t = 1, \ldots, (t_i - h)$ and evaluating the resulting model on t_i. We again use the MAE and the MSE as measures for the prediction accuracy.

The predictions of the null model can be used to calculate the following statistic for each of our models.

$$R^2_{\text{efron}}(f) = 1 - \frac{\sum_{i=t+1}^{t+h}(f(x_i) - y_i)^2}{\sum_{i=1}^{m}(\text{null}(x_i) - y_i)^2}, \tag{9.8}$$

where f is the model that maps predictors x_i to the response variable y_i, for which we calculate R^2_{efron}. This statistic is also known as Efron's pseudo R^2. It represents the improvement in terms of explained variation, of the prediction accuracy of the discovered model with respect to the null model. Negative values indicated that the predictive strength of the model f is worse than that of the null model.

9.1.4 Stress Identification

Besides the prediction of the exact NCO rates for each loan category, it is equally important to be able to predict stress from the macro-economic data. We say that a loan category P_i suffers from stress at time t if it exceeds a certain threshold T. We consider three thresholds for stress identification. We let threshold T_1 be the mean of the NCO series, T_2 the mean plus one times the standard deviation of the time-series, and T_3 the mean plus two times the standard deviation. This results in the following stress criteria,

$$
\begin{aligned}
T_1: &\quad \mathrm{NCO}_{i,t} > \overline{\mathrm{NCO}}_i, \\
T_2: &\quad \mathrm{NCO}_{i,t} > \overline{\mathrm{NCO}}_i + \mathrm{sd}(\mathrm{NCO}_i), \\
T_3: &\quad \mathrm{NCO}_{i,t} > \overline{\mathrm{NCO}}_i + 2\mathrm{sd}(\mathrm{NCO}_i).
\end{aligned}
$$

Of course, the mean and standard deviation are calculated based on the data in the training data only. For the predictions of a model f, we use the same threshold to determine whether or not the model predicts stress. Using this threshold, we can construct the number of observations correctly identified or true positives (TP), incorrectly identified or false positives (FP), correctly rejected or true negatives (TN), and incorrectly rejected or false negatives (FN) that the model f produces.

The prediction accuracy of stress identification can be evaluated by considering the following measures.

- Sensitivity or true positive rate is given by TP/(TP+FN) and specifies the ratio of positives that are correctly identified as such.
- Specificity or true negative rate is given by TN/(TN + FP) and specifies the ratio of negatives that are correctly identified as such.
- Accuracy is given by (TP + TN)(TP + FP + TN + FN) indicates the ratio of correctly categorized positives and negatives.
- Precision or the positive predictive values is given by TP/(TP+FP) and indicates the ratio of correctly identified positives with respect to the total number of identified positives.
- F1 is the harmonic mean of the sensitivity and the precision and is given by 2TP/(2TP + FP + FN).
- R^2_{count} is calculated as follows,

  $$R^2_{\text{count}} = \frac{\text{number correct} - f}{\text{total} - f}, \tag{9.9}$$

  where f is the count of the most frequent outcome. This measure can be seen as an adjusted version of the accuracy. By deducting the most frequent outcome from the resulting ratio, one corrects for the effect that a model which always predicts this outcome scores high on accuracy.

9.1.5 Residual Analysis

We further follow the suggested steps in Doornik & Hendry (2015) to test whether the automatically selected models are well-specified [51]. To check whether a model is well-specified, Doornik & Hendry (2015) asses whether the assumptions of linear regression hold[2]. These assumptions can be classified as follows[3].

1. The relation between the response variable y and the predictor variables is linear and additive. This means that,
   (a) the expected value of the response variable is a straight-line function of each predictor variable, holding all others fixed,
   (b) the slope of that line does not depend on the values of other variables,
   (c) the effects of different independent variables on the expected value of the dependent variable is additive.
2. The errors ϵ_t of the model are statistically independent.
3. The errors are homoskedastic,
   (a) versus time
   (b) versus predictions
   (c) versus any predictor variable
4. The errors are normally distributed. This assumption is actually not necessary for linear regression, it only ensures that the OLS method is the best linear unbiased estimator[4].

Linearity and additivity are tested by plotting observed versus predicted values. The data in this plot should be a straight line. The presence of heteroskedasticity can be detected by assessing plots of residuals versus predicted values and time versus residuals. We test for statistical independence of the errors by the Durbin-Watson test for serial autocorrelation, and analysis of the residual auto-correlation plot. The presence of arch-effects is tested by Engle's arch-test on the residuals. We test for normality of the error distribution with the Anderson-Darling test.

Note that a top-down stress testing model is primarily designed to determine how the NCO rates respond to macro-economic development. Modeling the complete behavior of the NCO rates provides aid in decomposing the time-series in a part determined by macro-economic developments, bank-characteristics, and (auto-regressive) noise.

[2] The assumptions in linear regression are discussed thoroughly in Appendix B.

[3] We follow the approach on Professor R.F. Nau's homepage, which can be found at: `http://people.duke.edu/~rnau/testing.htm`.

[4] For a more detailed description of the assumptions in linear regression, we refer to Appendix B.

9.2 Predictions with Individual Effects

For the evaluation of the models for loan categories P_i, $i = 1, \ldots, n_C$ and banks $b_j : j = 1, \ldots, n_B$, our main focus is again on the prediction accuracy. This is still primarily motivated by the objectives of stress testing. To adequately determine the feasibility and possible scope of a top-down model for stress-testing, our aim is to develop methods that accurately predict NCO rates conditional on exogenous paths of macro-economic variables. Prediction accuracy can be measured either in the cross-sectional or the time dimension. Since the top-down model is intended to predict future responses, our focus lies on the latter. For the training set, we consider the 100 largest observations of loan portfolios in a certain category in the 65 quarters of our training data. The remaining 31 quarters are used as a test set.

We determine the prediction accuracy as we did in Section 9.1 on the realized NCO rates for a small panel of banks. We are mainly interested in the possibilities of stress testing applied to individual banks' portfolios. To that end, we want to assess prediction accuracy for NCO rates of banks with large loan portfolios, comparable to the size of Rabobank's loan portfolio[5]. An advantage is that these large loan portfolios are expected to be less affected by random or idiosyncratic fluctuation in credit losses.

The panel consists of the 4 largest banks in the US, listed in Table 9.1 with the size of their loan portfolio[6]. For this panel of banks, we derive the out-of-sample statistics that were described in the previous Section9.1, to compare the performance of the models.

The second aspect of the individual model is to gain results on the ability of bank-specific CAMELS criteria to explain variability in NCO rates for individual bank loan portfolios. A panel data model is selected with the help of Adaptive Lasso. The inclusion or exclusion of bank-specific variables by the Adaptive Lasso method is in itself an indication of the relevance of these variables, since we know by Lemma 7.24 that the Lasso method selects the substantially relevant variables in a data-set. Therefore, we assess whether the CAMELS criteria explain variability in the cross-sectional dimension, by checking whether they are selected by Adaptive Lasso on the training set of the data.

[5] The current size of the credit portfolio of Rabobank is 434 billion euro [52].

[6] Morgan Stanley and Goldman Sachs were excluded from the validation panel, because they have only been commercial banks since 2008. Before they were so-called investment banks and hence, they did not traditionally have a large deposit and loan portfolio. Also, no bank data is available for those banks from 1991-2007.

Table 9.1: Bank panel used for Testing[1]

Bank Name	Total Loans[2]
JP Morgan Chase	546,788
Bank of America	746,142
Citigroup	343,161
Wells Fargo	797,284
Morgan Stanley	42,502
Goldman Sachs	35,219
HSBC NA	69,955
Bank of New York Mellon	17,535

[1] The amount of total loans was derived from the 2014 Q4 call report data.
[2] Reported values are in millions of dollars.

9.3 Top-Down Model for Individual Banks

When we have obtained models for all loan categories on each aggregation level, we aggregate the results to obtain prediction for individual banks.

For the evaluation of the individual portfolio model we consider the same test set of banks as we did for the models that include individual effects. We derive the same measures for the predictive strength of the model, as we did for the averaged models. In a stress testing exercise, it is typically assumed that all conditions remain equal, and only the macro-economic variables change according to the stress scenario. But when we make predictions for total NCO rates, the actual composition of the loan portfolio might change over period of the stress horizon. Therefore we recalculate the total NCO rate under the assumption that the composition does not change, to fairly compare the predicted rates with the realized rates.

Depending on the aggregation level the loan portfolio consists of a categories that are estimated. To recalculate the total predicted NCO rate at time $t+s$, under the loan composition at time $t+s$, we compute

$$\mathrm{NCO}_{t+s}^{\mathrm{tot}} = \sum_{i=1}^{a} \mathrm{NCO}_{i,t+s} \frac{\mathrm{TL}_{i,t+s}}{\mathrm{TL}_{t+s}^{\mathrm{tot}}}. \tag{9.10}$$

The results for different levels of aggregation are compared based on their prediction accuracy and stress identification on NCO rates for the loan portfolios of the bank listed in Table 9.1. Note that because we only have one sample of the macro-economy, and the observed charge-offs and recoveries for individual banks are subject to other processes that were not included in the model, that this method of testing the predictive strength of the model comes with some uncertainty. However, it may give us an indication of the reliability of the top-down model and to which extend it can be used for stress testing.

For the interpretation of all results we keep in mind the loan composition of the Big Four banks in the United States. From Table 9.2 we observe that Closed-End Residential Real Estate, Commercial & Industrial, Consumer, and Other are typically the largest loan categories.

Table 9.2: Composition of the Loan Portfolio for the Bank Panel

Loan Category	JP Morgan Chase	Bank of America	Citigroup	Wells Fargo
Home Equity Lines of Credit (HELOC)	10.7%	9.8%	5.3%	8.9%
Closed-End Residential Real Estate (RES)	25.8%	27%	25.8%	30.8%
Construction & Land Development (CLD)	0.9%	1.2%	0.4%	2.3%
Multi-Family (MF)	9.0%	0.7%	0.7%	1.5%
Non-farm Non-Residential (NFNR)	4.9%	6.0%	1.9%	11.4%
Farmland (FARM)	0.0%	0.2%	0.0%	0.3%
Commercial & Industrial (C&I)	16.2%	20.3%	10.5%	18.4%
Agriculture (AGRI)	0.1%	0.2%	0.1%	0.7%
Consumer (CON)	16.4%	22%	32.7%	13.4%
Leases on Financial Receivables (LEASE)	–	–	–	–
Depository Institutions (DEP)	1.0%	0.2%	2.8%	1.4%
Other (OTHER)	11.3%	7.2%	15.4%	7.6%

Part IV

Results

Chapter 10

Average NCO Rates

The top-down model for NCOs on the average loan portfolio of commercial banks in the United States was developed to assess the feasibility of top-down stress testing in general, and to determine whether Adaptive Lasso is able to discover an appropriate model.

Recall that the Adaptive Lasso model for an averaged loan category P_i is given by,

$$\mathrm{NCO}_{i,t}^{\mathrm{av}} = X_t \hat{\beta}^{\mathrm{alasso}} + \epsilon_{i,t},$$

where $\epsilon_{i,t}$ is a white noise process, and $\hat{\beta}^{\mathrm{alasso}}$ a sparse estimator of the coefficients of the linear model.

The RegARMA model is given by,

$$\begin{aligned}\mathrm{NCO}_{i,t}^{\mathrm{av}} &= X_t^{\mathrm{screen}} \hat{\beta} + \eta_{i,t} \\ \eta_t &= \sum_{j=1}^{p} \phi_{i,j} \eta_{i,t-j} + \sum_{j=1}^{q} \theta_{i,j} \epsilon_{i,t-j} + \epsilon_{i,t},\end{aligned}$$

where $\epsilon_{i,t}$ are i.i.d. random variables with mean zero, and X_t^{screen} consist of those columns of X for which $\hat{\beta}^{\mathrm{alasso}}$ has non-zero coefficients and p and q are chosen to minimize the BIC value.

Furthermore, recall that the results are benchmarked against a simple linear model, a null model, and a lagging model. We evaluate the results of the Adaptive Lasso, RegARMA, and linear benchmark model based on the economic interpretability of the automatically selected features in Section 10.1.1, the prediction accuracy based on out-of-sample criteria in Section 10.1.2, the ability to identify stress in Section 10.1.3, and the distribution of the residuals in Section 10.1.4. The results are presented for all loan categories at aggregation level A_1. From the comparison of the results for different loan categories we draw conclusions about minimal data requirements for top-down stress testing.

Recall the research questions that we posed in Chapter 3,

i *Can we use automated feature selection by employing Lasso methods to identify the model specification?*

ii *Can we construct a model that gives accurate predictions for NCO rates conditional on exogenous macro-economic paths?*

iii *What are the minimal data requirements for top-down modeling?*

iv *How do data limitations affect the prediction accuracy?*

v *Do we need to consider auto-regressive model specifications?*

In Section 10.1 we present the results and evaluation of the top-down models for average NCO rates per loan category. We use these results to answer the research questions in Section 10.2.

10.1 Model Results

In this section we present the results of the top-down models for average NCO rates of all commercial banks with domestic offices in the United States that we presented in Section 8.1.2. In Section 10.1 we evaluate the results of the average models in the way that was described in Section 9.1.

10.1.1 Selected Variables

The training set consists of the first 65 observations of NCO rates. The variables that were selected for the loan category Commercial & Industrial on the training set are presented in Table 10.1.

Table 10.1: Adaptive Lasso Selected Variables for NCO Rates on Commercial & Industrial

Selected Variables	Coefficient[1]
$\mathbb{1}_{\{\mathrm{GDP}_t<0\}}\mathrm{VIX}^2_{t-2}$	$+0.0462$
$\overline{\mathrm{DUR}}_t$	$+0.4766$
$\mathbb{1}_{\{\mathrm{VIX}_t>\mathrm{mean(VIX)}\}}\overline{\mathrm{HPI}}^2_{t-4}$	$+0.0920$
$\mathbb{1}_{\{\mathrm{CPPI}_t>0\}}\overline{\mathrm{UR}}^3_{t-4}$	-0.1060

[1] Coefficients are calculated based on normalized predictor variables and estimated on the training set.

It can be argued that the selected variables for the loan category Commercial & Industrial in Table 10.1 are economically significant. An interpretation can be as follows.

i If the GDP growth is negative, then the increase in NCO rates is proportional to the square of VIX, 2 quarters ago. The interpretation seems to be straightforward, since it suggests that as the economy is declining, the volatility on the stock markets in the previous quarters, is a measure for the severity of the increase in credit loss rates.

ii The average change in UR in the past year is proportional to the NCO rate. This variable is the most influential of the selected predictor variables, which corresponds to findings in literature, such as the CLASS model of Hirtle et al. [47].

iii If the level of VIX is above average, then the increase in NCO rates is proportional to the average squared change in HPI, one year ago. The significance of this predictor seem less unequivocal than the first two. A possible interpretation of the predictor $\mathbb{1}_{\{\text{VIX}_t > \text{mean(VIX)}\}} \overline{\text{HPI}}^2_{t-4}$ can be that the level of the change in house prices, is indicative of the severity of defaults and losses on Commercial & Industrial if the volatility on the stock markets is above average. A smaller change in house prices has less impact on the default rates. The estimated effect of this predictor variable is, however relatively small.

iv If CPPI is increasing, the change in NCO rates is negatively related to the yearly average UR in the previous year. The last predictor, $\mathbb{1}_{\{\text{CPPI}_t > 0\}} \overline{\text{UR}}^3_{t-4}$, with negative coefficient, could suggest that rising commercial property prices indicate a recovery of the economy leading to a decline in NCO rates proportional to past unemployment rates.

Of course, other explanations are possible. The above description is a mere indication of a possible interpretation of the selected variables. A thorough discussion of the potential meaning of the discovered predictor variables is well beyond the scope of this thesis, and requires solid economical background knowledge[1]. The variables that were selected by Adaptive Lasso for the other loan categories, can be found in Table 10.2.

[1] For the interested reader we recommend the book "Business Cycle Economics" by Todd Knoop [55] for an introduction to macro-economic theory.

Table 10.2: Selected variables on average NCOs

Loan Category	Selected Variables	Coefficient[2]
Loans to Depository Institutions	$\mathbb{1}_{\{\mathrm{CPPI}_t<0\}}\overline{\mathrm{INF}}_t^3$	+0.1105
Residential Real Estate[1]	$\mathbb{1}_{\{\mathrm{HPI}_t<0\}}\overline{\mathrm{HPI}}_{t-2}$	−0.3091
	$\mathbb{1}_{\{\mathrm{UR}_t>\mathrm{NROU}_t\}}\mathrm{UR}_t\overline{\mathrm{UR}}_{t-3}^3$	+0.0894
	$\mathbb{1}_{\{\mathrm{UR}_t>\mathrm{NROU}_t\}}\mathrm{UR}_t\overline{\mathrm{HPI}}_{t-4}$	−0.0339
	$\mathbb{1}_{\{\mathrm{HPI}_t<0\}}\mathrm{HPI}_t\overline{\mathrm{VIX}}_{t-4}^2$	−0.1019
HELOC[1]	$\mathbb{1}_{\{\mathrm{HPI}_t<0\}}\mathrm{DUR}_{t-2}$	−0.6080
	$\mathbb{1}_{\{\mathrm{HPI}_{t-1}<0\}}\overline{\mathrm{CPPI}}_{t-4}^2$	+0.2010
	$\mathbb{1}_{\{\mathrm{UR}_t>\mathrm{NROU}\}}\mathrm{UR}\overline{\mathrm{UR}}_{t-4}^2$	+0.2207
Construction and Land Development	$\mathbb{1}_{\{\mathrm{CPPI}_t<0\}}\mathrm{UR}_t^3$	+0.1909
	$\mathbb{1}_{\{\mathrm{CPPI}_t<0\}}\overline{\mathrm{INF}}_t$	+0.1430
	$\mathbb{1}_{\{\mathrm{CPPI}_t<0\}}\mathrm{CPPI}_t\overline{\mathrm{UR}}_t^3$	−1.557
	$\mathbb{1}_{\{\mathrm{CPPI}_{t-1}<0\}}\overline{\mathrm{INF}}_{t-2}^2$	+0.5004
Multi-Family Properties	$\mathbb{1}_{\{\mathrm{HPI}_t<0\}}\mathrm{UR}_t^3$	+0.1859
	$\mathbb{1}_{\{\mathrm{CPPI}_{t-1}<0\}}\overline{\mathrm{INF}}_{t-2}^2$	+0.2961
Nonfarm Nonresidential	$\mathbb{1}_{\{\mathrm{VIX}_t<\mathrm{mean(VIX)}\}}\mathrm{UR}_t^3$	+0.0510
	$\mathbb{1}_{\{\mathrm{CPPI}_t<0\}}\overline{\mathrm{UR}}_t^2$	+0.2083
	$\mathbb{1}_{\{\mathrm{CPPI}_t<0\}}\mathrm{CPPI}_t\overline{\mathrm{UR}}_t^3$	−0.0350
	$\mathbb{1}_{\{\mathrm{CPPI}_{t-1}<0\}}\overline{\mathrm{INF}}_{t-2}^2$	+0.1805
Farmland[1]	UR_t^3	+0.0689
Commercial & Industrial	$\mathbb{1}_{\{\mathrm{GDP}<0\}}\mathrm{VIX}_{t-2}^2$	+0.0462
	$\overline{\mathrm{DUR}}_t$	+0.4766
	$\mathbb{1}_{\{\mathrm{VIX}_t>\mathrm{mean(VIX)}\}}\overline{\mathrm{HPI}}_{t-4}^2$	+0.0920
	$\mathbb{1}_{\{\mathrm{CPPI}_t>0\}}\overline{\mathrm{UR}}_{t-4}^3$	−0.1060
Consumer Loans[3]	$\mathbb{1}_{\{\mathrm{DUR}_{t-1}>0\}}\mathrm{INF}_{t-1}$	−0.0040
	$\mathbb{1}_{\{\mathrm{VIX}_t>\mathrm{mean(VIX)}\}}\overline{\mathrm{INF}}_{t-4}^2$	+0.1881
	$\mathbb{1}_{\{\mathrm{CPPI}_t<0\}}\overline{\mathrm{DUR}}_t$	+0.1992
	$\mathbb{1}_{\{\mathrm{DUR}_t<0\}}\mathrm{DUR}_t\overline{\mathrm{INF}}_t^3$	+0.1011
	$\mathbb{1}_{\{\mathrm{DI}_{t-1}<0\}}\overline{\mathrm{DUR}}_{t-2}$	+0.0438
	$\mathbb{1}_{\{\mathrm{DJI}_t<0\}}\mathrm{DJI}_t\overline{\mathrm{DJI}}_{t-3}^2$	+0.0134
	$\mathbb{1}_{\{\mathrm{UR}_t>\mathrm{NROU}_t\}}\overline{\mathrm{CPPI}}_{t-2}$	+0.0207
Agricultural Loans	$\mathbb{1}_{\{\mathrm{VIX}_{t-1}>\mathrm{MEAN(VIX)}\}}\overline{\mathrm{HPI}}_{t-2}^3$	+0.0976
	$\mathbb{1}_{\{\mathrm{DUR}_t>0\}}\overline{\mathrm{CPPI}}_{t-4}^3$	+0.0572
Leases and Financing Receivables	$\mathbb{1}_{\{\mathrm{HPI}_t>0\}}\mathrm{HPI}_t\mathrm{VIX}_{t-1}$	+1.174

Loan Category	Selected Variables	Coefficient[2]
	$\overline{\text{DJI}}_t\text{UR}_{t-1}$	-0.1796
	$\overline{\text{VIX}}_t$	$+0.0795$
	$\overline{\text{UR}}_{t-4}$	-0.1252
	$\mathbb{1}_{\{\text{DI}_t>0\}}\overline{\text{DUR}}_t$	$+0.2741$
	$\mathbb{1}_{\{\text{GDP}_{t-1}<0\}}\overline{\text{VIX}}_t^3$	$+1.380$
	$\mathbb{1}_{\{\text{CPPI}_{t-1}<0\}}\overline{\text{DUR}}_{t-2}$	$+0.9257$
	$\mathbb{1}_{\{\text{DJI}_t<0\}}\text{DJI}_t\overline{\text{INF}}_{t-4}$	-0.1812
	$\mathbb{1}_{\{\text{CPPI}_t<0\}}\overline{\text{HPI}}_{t-4}^2$	$+1.692$
	$\mathbb{1}_{\{\text{DI}_t<0\}}\text{DI}_t\overline{\text{HPI}}_{t-4}^3$	$+0.0230$
Other Loans	$\mathbb{1}_{\{\text{HPI}_t<0\}}\text{HPI}_t\overline{\text{INF}}_t$	-0.1218
	$\mathbb{1}_{\{\text{DUR}_t>0\}}\overline{\text{INF}}_{t-3}^2$	$+0.0889$
	$\mathbb{1}_{\{\text{VIX}_t>\text{mean(VIX)}\}}\text{VIX}_t\overline{\text{INF}}_{t-4}^2$	$+0.0604$
Commercial Real Estate Loans	$\mathbb{1}_{\{\text{VIX}_t<\text{mean(VIX)}\}}\text{UR}_t^3$	$+0.2201$
	$\mathbb{1}_{\{\text{CPPI}_t<0\}}\overline{\text{UR}}_t^3$	$+0.8435$
	$\mathbb{1}_{\{\text{CPPI}_t<0\}}\text{CPPI}_t\overline{\text{UR}}_t^3$	-0.8802
	$\mathbb{1}_{\{\text{CPPI}_{t-1}<0\}}\overline{\text{INF}}_{t-2}^2$	$+0.4755$
	$\mathbb{1}_{\{\text{HPI}_t<0\}}\overline{\text{VIX}}_{t-4}^3$	-0.9105
Residential Real Estate Loans[1]	$\mathbb{1}_{\{\text{HPI}_{t-1}<0\}}\overline{\text{HPI}}_{t-2}$	-0.4315
	$\mathbb{1}_{\{\text{HPI}_t<0\}}\text{HPI}_t\overline{\text{VIX}}_{t-4}$	-0.0392
	$\mathbb{1}_{\{\text{UR}_t>\text{NROU}_t\}}\text{UR}_t\overline{\text{HPI}}_{t-4}$	-0.0305
	$\mathbb{1}_{\{\text{HPI}_{t-1}<0\}}\overline{\text{CPPI}}_{t-4}^2$	$+0.0340$
	$\mathbb{1}_{\{\text{UR}_t>\text{NROU}_t\}}\text{UR}_t\overline{\text{UR}}_{t-4}^3$	-0.1262
Real Estate Loans	$\mathbb{1}_{\{\text{CPPI}_{t-1}<0\}}\overline{\text{INF}}_{t-2}^2$	$+0.1995$
	$\mathbb{1}_{\{\text{HPI}_t<0\}}\overline{\text{VIX}}_{t-4}^3$	$+0.1681$
Total Loans[1]	$\overline{\text{DUR}}_t$	$+0.1289$
	$\mathbb{1}_{\{\text{HPI}_t<0\}}\text{HPI}_t\overline{\text{UR}}_{t-2}$	-0.1829
	$\mathbb{1}_{\{\text{HPI}_t<0\}}\overline{\text{VIX}}_{t-3}^2$	-0.3156
	$\mathbb{1}_{\{\text{HPI}_t<0\}}\overline{\text{DUR}}_{t-4}$	$+0.0194$

[1] Variables were selected on the full sample (training and test data combined).
[2] Coefficients were calculated based on normalized predictor variables.
[3] Series were detrended prior to the Adaptive Lasso procedure.

We shall not discuss the interpretation of the selected model specifications for all loan categories here. But we do note that for NCO rates on the categories Closed-End Residential Real Estate (RES) and Home Equity Lines of Credit (HELOC) no variables were selected on the training set. From the loan composition of the

Big Four banks in our test panel of banks in Table 9.2, we can observe that these categories make up a large part of a typical loan portfolio. Therefore, it is surely undesirable not to have a top-down model for NCOs on these loan categories. To understand these results better we therefore present the selected model specification on the full sample for loan categories that are not modeled well. For HELOC and RES it turns out that the most relevant predictors, with the highest normalized coefficients, involve interactions and transformations of HPI. From Figure 5.2 it is clear that the house price index follows an extreme path during the time of our test sample. In particular, HPI was negative only once in our training sample, which goes from 1991 Q1 to 2006 Q2. It is therefore not surprising that Adaptive Lasso cannot pick up these trends from the training data alone.

It is remarkable that Adaptive Lasso is able to select parsimonious models for NCO rates on different loan categories from a set of more than 10000 complex interactions and transformations of predictors. Surely, it is fascinating to observe that sensible, economically interpretable and simple linear models can be obtained automatically. These hypothesis generated by Adaptive Lasso are such that a modeler can pass expert judgment to the selected models and can make adjustments to them if it is needed. In the next section we shall see that these models can also give accurate predictions in a stress scenario.

10.1.2 Prediction Accuracy

As discussed in Chapter 9 we measure prediction accuracy by determining the Mean Square Error (MSE), Mean Absolute Error (MAE), and R^2_{efron} of the predictions. The latter is a measure for the amount of variability in the predicted values that is explained by a model, as compared to the null model. The predictive strength of the models can be understood intuitively by analyzing plots of observed versus predicted values. Therefore, we also show the actual, fitted, and predicted values in a plot. With actual values we refer to the NCO rate time-series that are filtered according to equation 6.2. Lastly, we look at the results for longer stress horizons.

10.1.2.1 Prediction Plots

In Figures 10.1, 10.2, 10.3, 10.4, 10.5, and 10.6 a comparative plot for the Adaptive Lasso, RegARMA, and linear benchmark model is displayed for the 9-quarters-ahead predictions of NCO rates on all loan categories that are in the lowest aggregation level A_1. Recall that the 9-quarters-ahead predictions are obtained by applying an extending window. That is, the dotted line in these figures represents the predicted values based on a model estimated on training data up to 9 quarters before the time of prediction. The results will be discussed for each loan category separately.

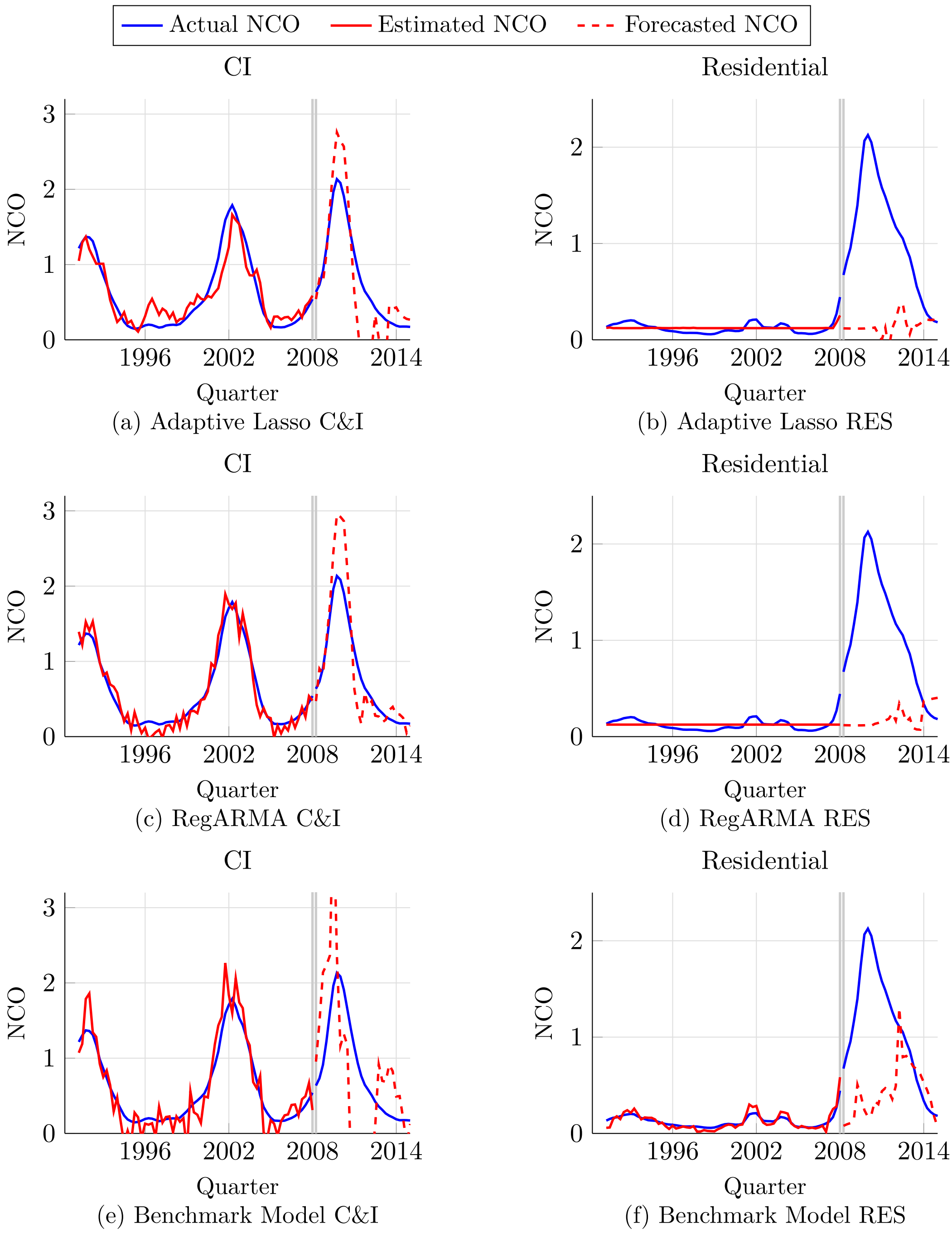


Figure 10.1: 9 quarters ahead forecasts for C&I and RES Loans

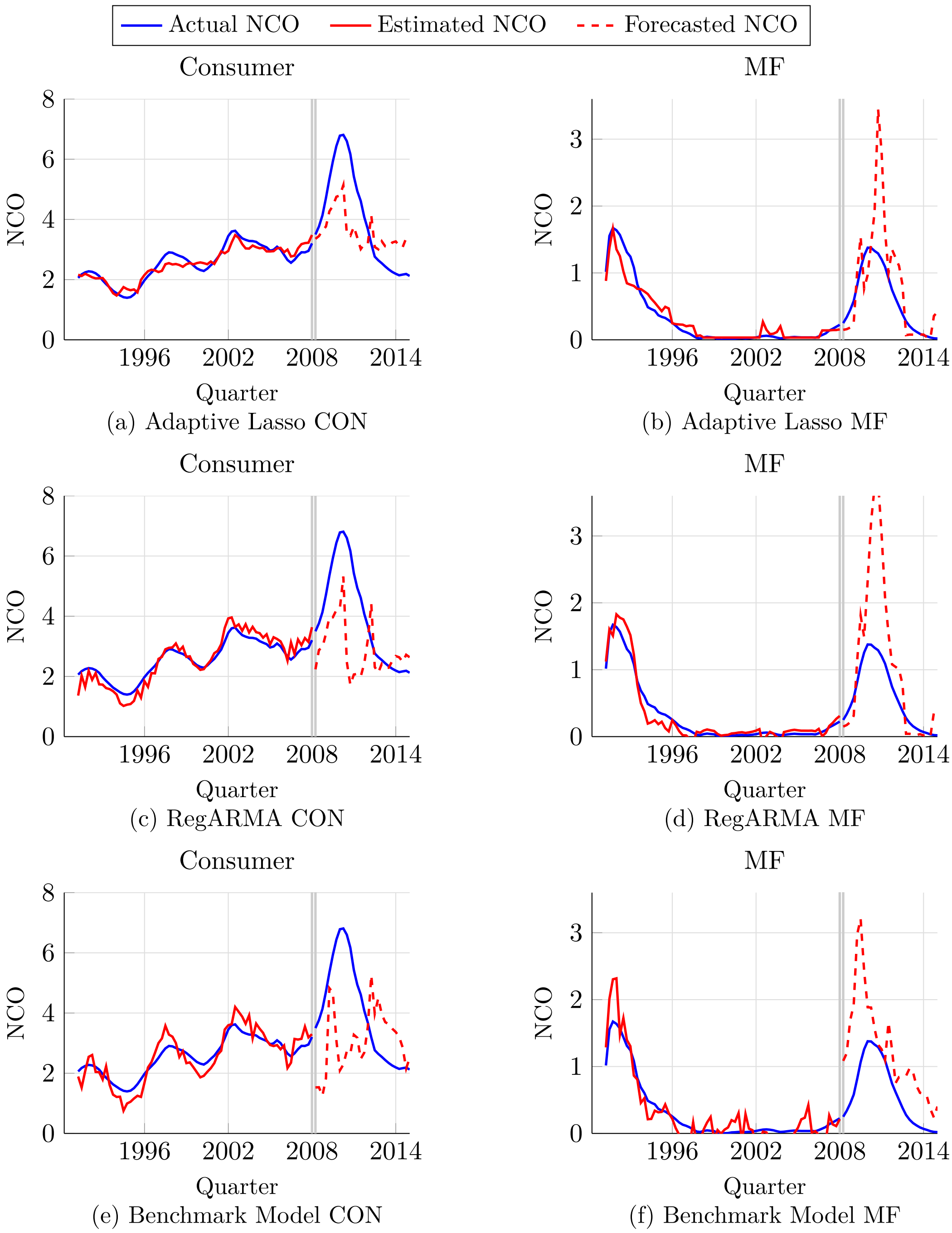
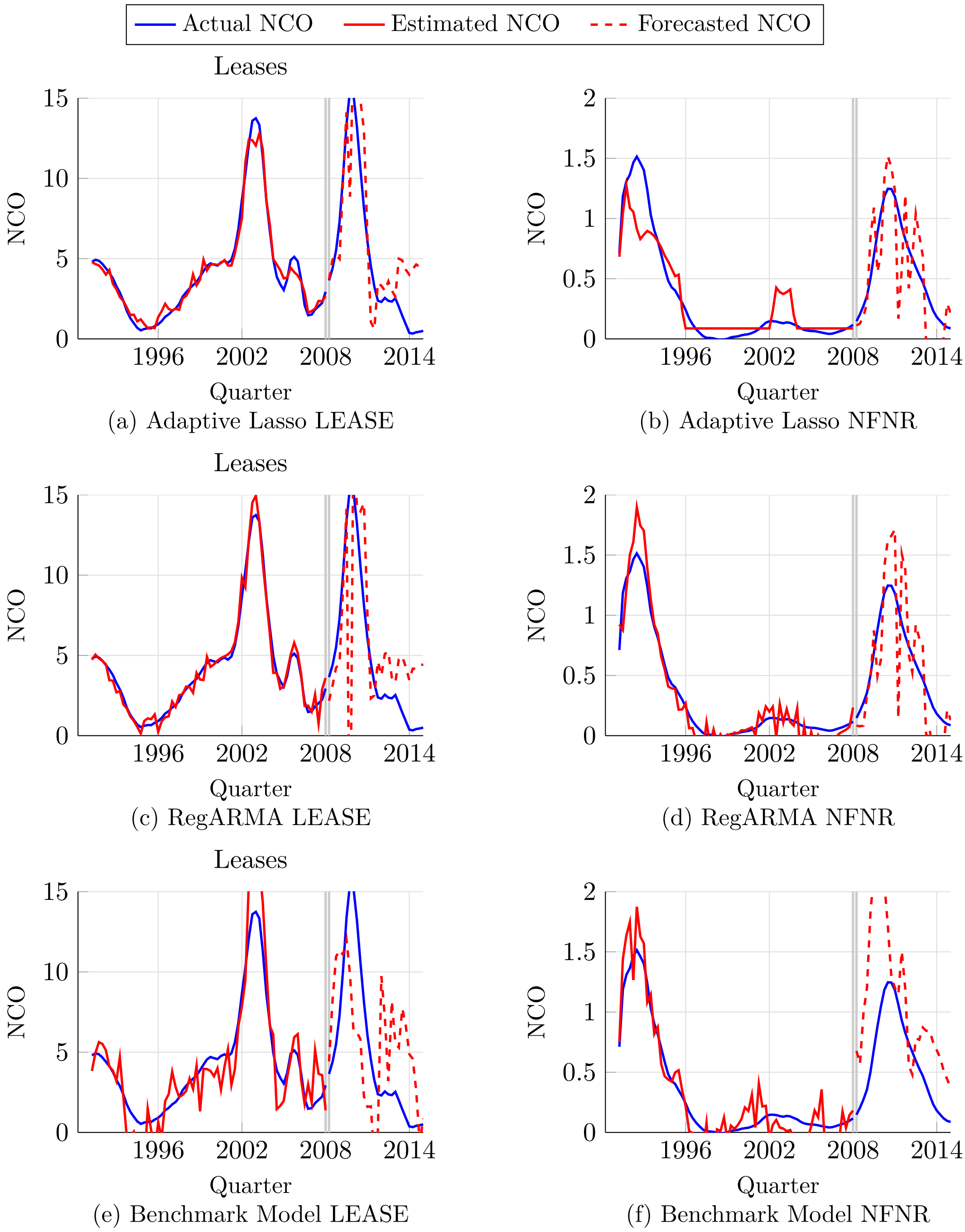


Figure 10.2: 9 quarters ahead forecasts for CON and CLD Loans



(a) Adaptive Lasso LEASE

(b) Adaptive Lasso NFNR

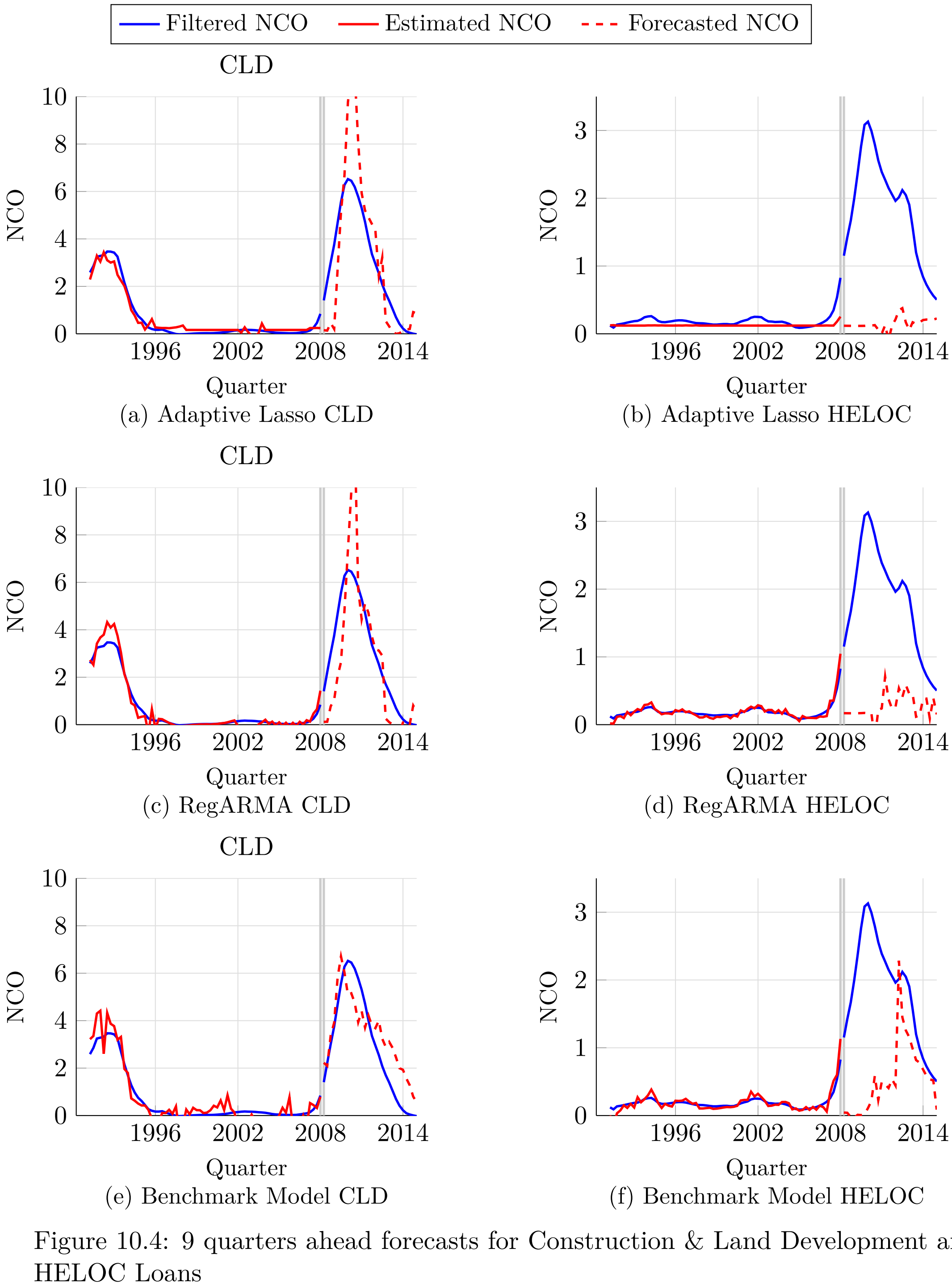
(c) RegARMA LEASE

(d) RegARMA NFNR

(e) Benchmark Model LEASE

(f) Benchmark Model NFNR

Figure 10.3: 9 quarters ahead forecasts for LEASE and NFNR.



(a) Adaptive Lasso CLD

(b) Adaptive Lasso HELOC

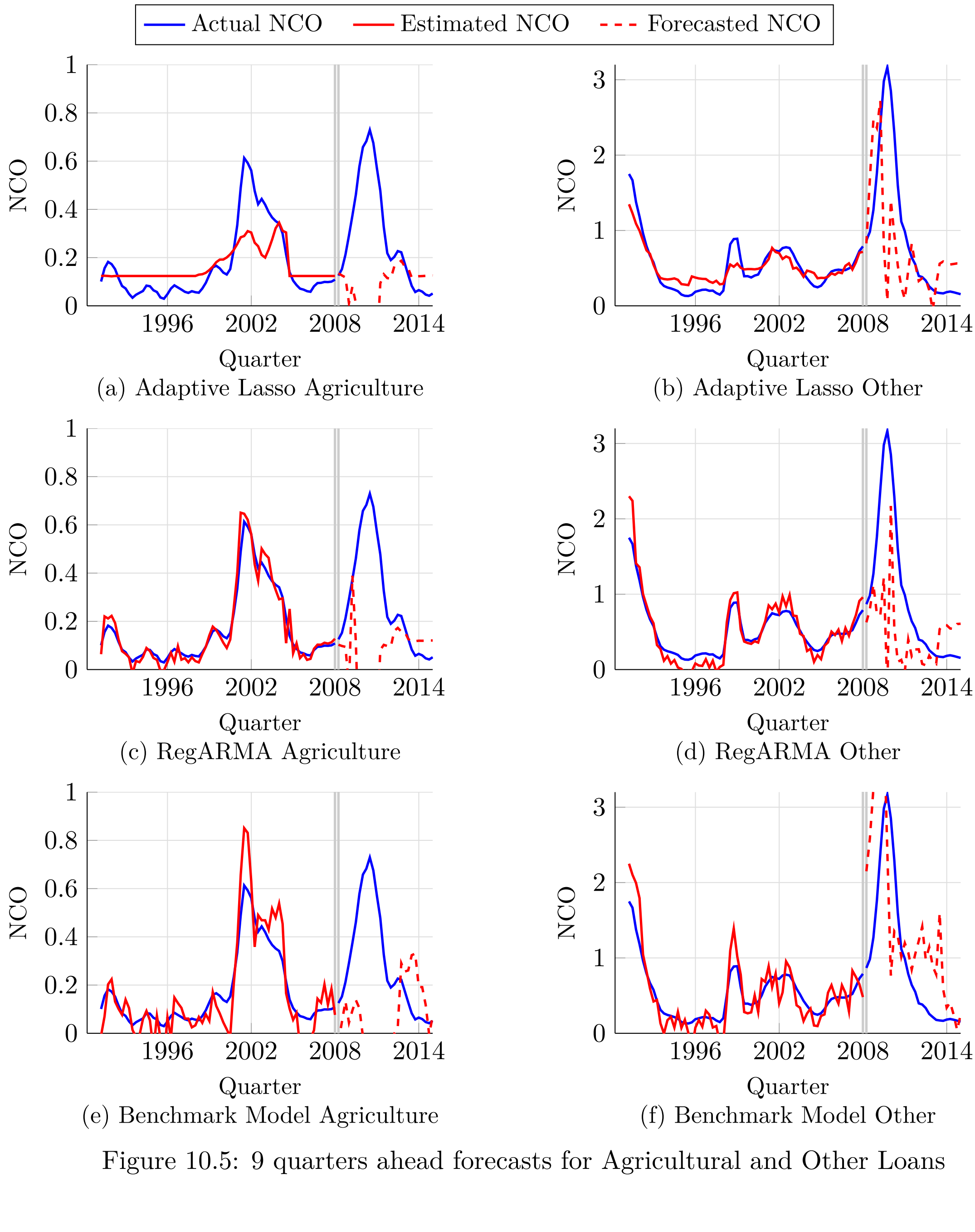
(c) RegARMA CLD

(d) RegARMA HELOC

(e) Benchmark Model CLD

(f) Benchmark Model HELOC

Figure 10.4: 9 quarters ahead forecasts for Construction & Land Development and HELOC Loans



(a) Adaptive Lasso Agriculture

(b) Adaptive Lasso Other

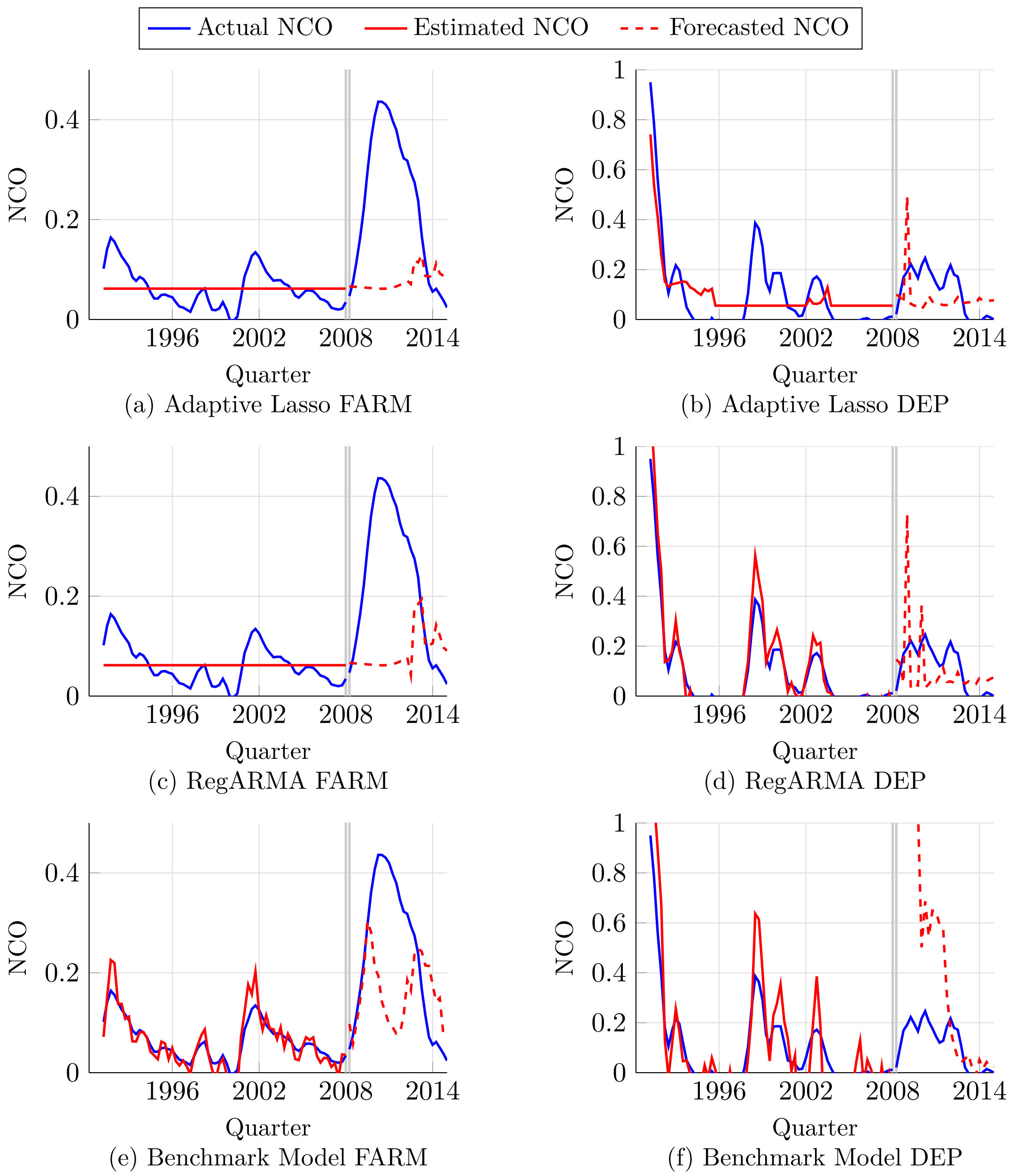
(c) RegARMA Agriculture

(d) RegARMA Other

(e) Benchmark Model Agriculture

(f) Benchmark Model Other

Figure 10.5: 9 quarters ahead forecasts for Agricultural and Other Loans



(a) Adaptive Lasso FARM
(b) Adaptive Lasso DEP
(c) RegARMA FARM
(d) RegARMA DEP
(e) Benchmark Model FARM
(f) Benchmark Model DEP

Figure 10.6: 9 quarters ahead forecasts for FARM and DEP Loans.

For NCO rates on Commercial & Industrial in Figures 10.1a, 10.1c, and 10.1e the 9-quarters ahead predictions starting from 2008 Q1 resemble the actual NCO rates best for the RegARMA model. Although the fit for the Adaptive Lasso model is slightly worse for low rates, it captures the peak around 2009 very adequately. The 9-quarters-ahead predictions of the linear benchmark model do not seem to capture the behavior of the NCO rates well.

None of the models capture the behavior of NCO rates on Closed-End Residential Real Estate (RES) loans well, as can be seen from Figures 10.1b, 10.1d, and 10.1f. We remark that the results for the linear benchmark model are better than those of the Adaptive Lasso and RegARMA models, but they are not adequate. As we mentioned in our discussion of the selected variables for RES this is not surprising, because the training data for this loan category is too limited. It cannot be expected that a completely data-driven method will capture this effect on such a training sample.

In Figure 10.2 we observe that NCO rates on Consumer (CON) loans there is a linear time-trend. In the estimation procedure this was taken into account. However, especially the drop in the predictions of the RegARMA and benchmark model indicate that this issue needs to be addressed further. We do note that the benchmark model is the only model that is incapable of capturing the peak in NCO rates for CON loans. For Multi-Family (MF) NCO rates all models overshoot the peak in rates. Adaptive Lasso seems to have the better overall fit.

The results for LEASE in Figure 10.3 show that Adaptive Lasso captures the NCO rate best, and the linear benchmark has the worst performance. For Non-farm Non-Residential (NFNR) the differences are difficult to determine from the plot. The 9-quarter ahead predictions for Construction & Land Development (CLD) in Figure 10.4 indicate that the NCO rate is predicted accurately by all models, but the fit is better for the benchmark model, because the Adaptive Lasso and RegARMA model overshoot the peak in rates. The predictions for Home Equity Lines of Credit (HELOC) loans have a similar interpretation as those for Closed-End Residential Real Estate (RES) loans. This can be explained in the same fashion.

None of the models gives accurate prediction for the loan categories Agriculture (AGRI), Farmland (FARM), and Depository Institutions (DEP) in Figures 10.6 and 10.5. This can be explained by more irregular behavior of the NCO rates on these loan categories. Fortunately, these categories comprise only a small part of a typical loan portfolio, as can be seen from Table 9.2. Lastly, the 9-quarter-ahead predictions for NCO rates on Other (OTHER) loans is not optimal for any of the models. However, the Adaptive Lasso and benchmark model give quite adequate results, taking into account the irregular behavior of the NCO rates in the training set.

Overall, we conclude that the parsimonious models selected and estimated by Adaptive Lasso give accurate predictions, provided that the scenario in the test set is not too dissimilar from the training data, and regularities in the training data can be discovered. Furthermore, RegARMA gives very reasonable results, despite the fact that it was estimated in two steps. Both the Adaptive Lasso and RegARMA models appear to have better overall performance than the benchmark model. But even the prediction accuracy of the benchmark model suggests that NCO rates are predictable from macro-economic variables. We note that where Adaptive Lasso has the smallest prediction error, the fitted values on the training set appear to be better for the RegARMA and benchmark models.

10.1.2.2 Prediction Error Measures

To draw conclusions about the prediction accuracy of top-down models for stress testing, we take a closer look at some of the measures we considered in Section 9.1.3. In Table 10.3 the prediction error, in terms of MSE and MAE, and R^2_{efron} of each model is compared on 9-quarters-ahead predictions for all loan categories. As an extra column the error for the *null* model is calculated.

Table 10.3: Prediction Errors 9 quarters ahead

	Adaptive Lasso			RegARMA			Benchmark			Null	
	MSE	MAE	R^2_{efron}	MSE	MAE	R^2_{efron}	MSE	MAE	R^2_{efron}	MSE	MAE
HELOC	3.33	1.63	−0.08	3.14	1.59	−0.02	2.74	1.35	0.11	3.09	1.51
RES	1.27	0.94	−0.06	1.23	0.95	−0.03	0.89	0.73	0.25	1.20	0.91
CLD	3.62	1.41	0.64	3.32	1.29	0.67	1.41	1.06	0.86	10.13	2.58
MF	0.37	0.39	−0.12	0.82	0.58	−1.48	0.72	0.66	−1.17	0.33	0.47
NFNR	0.09	0.25	0.59	0.11	0.27	0.50	0.42	0.53	−0.85	0.23	0.38
C&I	0.16	0.33	0.69	0.15	0.27	0.71	0.67	0.66	−0.32	0.51	0.57
LEASE	8.97	2.41	0.63	16.56	2.98	0.33	20.47	3.94	0.17	24.54	3.97
CON	1.76	1.14	0.63	3.35	1.39	0.30	5.06	1.89	−0.05	4.82	1.76
OTHER	0.88	0.62	0.25	0.95	0.68	0.19	0.92	0.76	0.22	1.17	0.81
AGRI	0.11	0.23	−0.69	0.11	0.22	−0.62	0.12	0.28	−0.83	0.07	0.19
FARM	0.05	0.18	0.02	0.05	0.18	0.02	0.03	0.13	0.42	0.05	0.18
DEP	0.01	0.10	−0.51	0.02	0.11	−1.40	0.61	0.50	−66.75	0.01	0.09

The results seem to favor the Adaptive Lasso method. We shall elaborate on this further, but first we confine our discussion. First off, we shall exclude the categories Multi-Family, Agriculture and Depository Institutions from our discussion, because for these loan categories the null model outperforms all models under consideration. In Figures 10.2b, 10.2d, and 10.2f, we observe that all methods tend to overshoot the peak in NCO rates for MF loans, which leads to a relatively high MSE and even leads to a negative R^2_{efron}. Indicating that the null model explains more of the variability in the NCO rate than all the other models. Note that especially Agriculture and Depository Institutions display irregular behavior and constitute only a small part of the loan composition of the banks in our test panel by the results in Table 9.2.

Second, we exclude the categories Home Equity Lines of Credit and Closed-End Residential Real Estate from our discussion. Since the values for the home price index during the recession period in our test data is far beyond the range of this index in our training data, none of the models is able to capture the response of the NCO rates to a satisfactory degree. Furthermore, we remark that from Table 10.2 we can observe that transformations of the home price index are selected on the full sample (training and test sample combined). We also remark here that, for these loan categories, a 9-quarter lagging model, outperforms all models under consideration, including the null model.

For the remaining categories, we see that Adaptive Lasso gives the most accurate predictions for the loan categories Non-farm Non-Residential, Consumer, Leases on Financial Receivables, and Other with an R^2_{efron} of 0.59, 0.63, 0.63, and 0.25 respectively. Note that this measures indicates the percentage of variability in the data is explained by the model, as compared to the null model. RegARMA has an R^2_{efron} value of 0.71 for Commercial & Industrial, but is closely followed by Adaptive Lasso with 0.69. The benchmark model explains more of the variation in the data compared to the null model for the categories Farmland and Construction & Land Development, with an R^2_{efron} of 0.42 and 0.86. Note that for the latter category both Adaptive Lasso and RegARMA have satisfactory power with R^2_{efron} of 0.64 and 0.67 respectively.

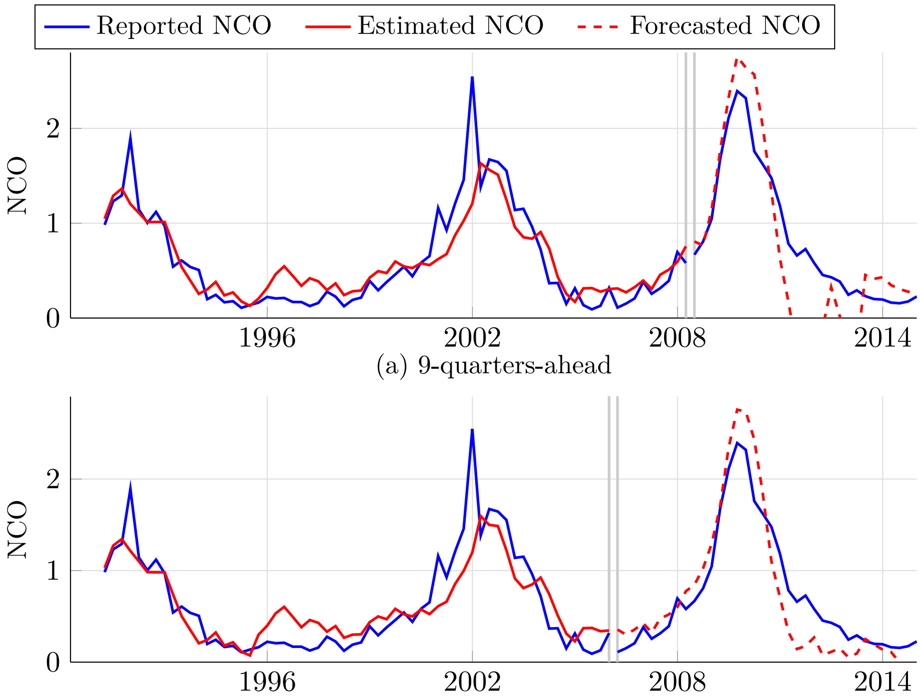
Taking into account that for most banks the loan categories Closed-End Residential Real Estate, Commercial & Industrial, Non-farm Non-Residential, and Consumer are the largest components of the loan portfolio by the results in Table 9.2, we conclude that Adaptive Lasso is the best model to predict NCO rates on the average loan portfolio of commercial banks in the United States conditional on macro-economic paths 9-quarters-ahead, for the period 2007-2014. We note here that Adaptive Lasso is closely followed by the RegARMA model in terms of predictive strength.

10.1.2.3 Longer Stress Horizons

The next step is to evaluate our method for predictions of more than 9 quarters ahead. Therefore, we also evaluate the performance of our models on the complete test set, which consists of averaged NCO rate data from 2007 Q1 to 2014 Q4. This represents the purposes of stress testing most closely.

In Figures 10.7a and 10.7b, we compare the 9-quarters-ahead predictions of the Adaptive Lasso model on the last 31 quarters (nearly 8 years) of our data-set. The predicted values are based on a model that was estimated on the training set, which consists of the first 65 quarters of our data-set. We conclude that the results for

9-quarters-ahead forecasts and the results on a test set of nearly 8 years have comparable predictive strength for the loan category Commercial & Industrial during the 2007-2009 crisis and the recovery that followed. Actually, since the dotted line in Figure 10.7b is closer to the reported values than the dotted line in Figure 10.7a, the prediction accuracy on the test set is slightly better than the 9-quarters-ahead predictions. This suggests that, with Adaptive Lasso, it is possible to extend the stress horizon far beyond 9 quarters.



(a) 9-quarters-ahead

(b) Test Set

Figure 10.7: Comparison of Adaptive Lasso Results for Commercial & Industrial for 9-quarters-ahead Predictions and the Test Set.

Table 10.4 shows the prediction accuracy of all models on the test set. The results for all loan categories on the test set are very similar to those for the 9-quarter-ahead forecasts in Table 10.3. With the exception of the categories Multi-Family and Non-farm Non-Residential, which have much higher and lower prediction errors respectively. Remarkably, the predictive strength of Adaptive Lasso does not decrease when the stress horizon is extended. On the stress test set, Adaptive Lasso remains the model with the most accurate predictions, followed by the RegARMA model. Although these accuracies were obtained from only one sample, the results indicate that stress horizons up to five years are feasible in top-down stress testing.

Table 10.4: Prediction Errors on the Stress Test Set

	Adaptive Lasso			RegARMA			Benchmark			Null	
	MSE	MAE	R^2_{efron}	MSE	MAE	R^2_{efron}	MSE	MAE	R^2_{efron}	MSE	MAE
HELOC	3.13	1.56	0.00	3.12	1.56	0.00	3.41	1.64	−0.09	3.12	1.56
RES	1.17	0.89	0.00	1.17	0.89	0.00	1.03	0.85	0.11	1.17	0.89
CLD	5.24	1.69	0.43	7.42	2.17	0.19	1.30	0.98	0.86	9.18	2.29
MF	0.08	0.22	0.71	0.08	0.22	0.73	0.69	0.70	−1.44	0.28	0.40
NFNR	0.20	0.31	0.03	0.32	0.45	−0.62	0.40	0.50	−0.98	0.20	0.35
C&I	0.10	0.23	0.77	0.16	0.27	0.63	0.81	0.57	−0.89	0.43	0.49
LEASE	7.87	1.96	0.62	10.48	2.09	0.50	11.62	2.80	0.44	20.76	3.50
CON	1.49	1.02	0.66	4.29	1.59	0.01	4.11	1.55	0.05	4.34	1.52
OTHER	0.75	0.56	0.26	0.96	0.65	0.04	0.62	0.64	0.38	1.00	0.67
AGRI	0.10	0.21	−0.69	0.13	0.25	−1.17	0.11	0.24	−0.86	0.06	0.17
FARM	0.05	0.16	0.00	0.05	0.17	0.00	0.03	0.12	0.45	0.05	0.17
DEP	0.01	0.10	−0.53	0.01	0.09	−0.31	0.54	0.55	−62.85	0.01	0.08

10.1.3 Stress Identification

In this section we test whether our method is capable of identifying stress events from non-stress events. To that end, recall that we defined stress as follows. An averaged loan category P_i is *stressed* at time t when,

$$\text{NCO}^{\text{av}}_{i,t} \geq \text{mean}(\text{NCO}^{\text{av}}_{i}) + i \cdot \text{sd}(\text{NCO}^{\text{av}}_{i}), \quad , i = 0, 1, 2. \tag{10.1}$$

We thus consider three thresholds T_1, T_2, and T_3. Using this definition we can determine whether the models can distinguish between stress and no stress for different thresholding values. In Table 10.5 below we show the results for the loan category Commercial & Industrial for forecasts on the stress test set. Adaptive Lasso and the RegARMA model score better on all thresholds for all criteria than the benchmark model and null model. The RegARMA model produces slightly better results than the Adaptive Lasso model. The high amount of true positives and true negatives for Adaptive Lasso we conclude that the top down model is able to forecast stress to a high degree of accuracy on the test sample.

It is interesting to calculate the same scores on the 9-quarter-ahead predictions. In Table 10.6 the results for the category Commercial & Industrial are displayed. A key observation is that the specificity and sensitivity scores are remarkably high for both the forecast on the test set and the 9-quarter ahead predictions, indicating that it is indeed possible to use a top-down model to stress NCO rates on Commercial & Industrial loans.

For the results of the specificity, sensitivity, accuracy, and R^2_{count} for all other loan categories we refer to Appendix E. In Table 10.7 the F1 scores for all averaged loan categories predicted on the test set are displayed. Note that the zero scores for the F1 score at the third thresholding level for Non-farm Non-Residential loans is caused by

Table 10.5: Stress Identification Results for 9-quarters-ahead Predictions of C&I NCO

	Adaptive Lasso			RegARMA			Benchmark			Null		
	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
Specificity	0.92	0.92	0.92	0.94	1	0.93	0.83	0.83	0.83	1	1	1
Sensitivity	1	1	1	0.79	1	1	0.5	0.5	0.5	0	0	0
Accuracy	0.93	0.93	0.93	0.873	1	0.94	0.79	0.79	0.79	0.86	0.86	0.86
R^2_{count}	0.71	0.88	0.67	0.71	1.00	0.67	0.42	0.70	0.29			
F1	0.8	0.8	0.8	0.85	1	0.8	0.4	0.4	0.4	0	0	0
TP	10	7	6	11	8	4	12	7	2	0	0	0
FP	1	0	0	1	0	2	6	4	4	0	0	0
TN	16	17	17	16	23	25	11	19	23	17	23	27
FN	4	7	8	3	0	0	2	1	2	14	8	4

Table 10.6: Stress Identification Results for Forecasting C&I NCOs on the Test Set

	Adaptive Lasso			RegARMA			Benchmark			Null		
	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
Specificity	1	1	0.92	1	1	0.92	0.57	0.85	0.83	0.07	1	1
Sensitivity	0.71	0.88	1	0.71	1	1	0.64	0.75	0.5	0.07	0	0
Accuracy	0.86	0.96	0.93	0.86	1	0.93	0.61	0.82	0.79	0.07	0.71	0.86
R^2_{count}	0.71	0.88	0.50	0.71	1.00	0.50	0.21	0.38	0.00			
F1	0.83	0.93	0.8	0.83	1	0.8	0.62	0.71	0.4	0.07	0	0
TP	10	7	4	10	8	4	9	6	2	1	0	0
FP	0	0	2	0	0	2	6	3	4	13	0	0
TN	14	20	22	14	20	22	8	17	20	1	20	24
FN	4	1	0	4	0	0	5	2	2	13	8	4

the fact that there are no observations of stress levels above this threshold. We also note that the categories Home Equity Lines of Credit and Closed-End Residential Real Estate are stressed at each time point for all thresholds, due to the extremity of the scenario for the house price index on the test sample.

A striking result is that where the benchmark model outperformed Adaptive Lasso on the loan category Construction & Land Development based on MSE, here Adaptive Lasso actually seems to be a better indicator for stress for thresholds T_1 and T_2. Since Adaptive Lasso has the highest F1 score for most categories and has the highest average F1 score, we conclude that it predicts stress most accurately on our test set.

10.1.4 Residual Analysis

In this section we look at the residuals of the estimation on the training set for the loan category Commercial & Industrial, for the Adaptive Lasso, RegARMA, and benchmark models. In the Figures 10.8, 10.9, and 10.10 the residuals analysis for the

Table 10.7: F1 Score for Stress Identification by Category on the Test Set

	Adaptive Lasso			RegARMA			Benchmark		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
HELOC	0.81	0.53	0.35	0.78	0.67	0.53	0.81	0.81	0.76
RES	0.73	0.56	0.14	0.76	0.56	0.35	0.92	0.90	0.87
CLD	0.76	0.90	0.83	0.84	0.90	0.78	0.88	0.83	0.83
MF	0.85	0.73	0.40	0.85	0.82	0.50	0.77	0.62	0.46
NFNR	0.91	0.56	0.00	0.91	0.71	0.00	0.78	0.72	0.00
C&I	0.83	0.93	0.80	0.83	1.00	0.80	0.62	0.82	0.44
LEASE	0.74	1.00	0.67	0.64	0.92	0.67	0.53	0.38	0.29
CON	0.83	0.84	0.57	0.59	0.64	0.50	0.67	0.39	0.20
OTHER	0.67	0.67	0.62	0.53	0.40	0.22	0.79	0.71	0.70
AGRI	0.21	0.00	0.00	0.11	0.20	0.00	0.08	0.00	0.00
FARM	0.54	0.25	0.00	0.50	0.24	0.29	0.86	0.77	0.67
DEP	0.12	0.00		0.29	0.00	0.00	0.88	0.00	0.00

adaptive Lasso, RegARMA, and the benchmark model respectively, are displayed.

For all models, the observed vs predicted plots show a strong linear dependency, as is required for linear models. It must be noted that the variance for the Adaptive Lasso model fit is notably larger. The residuals vs predicted plots show some odd behavior for predicted values close to zero. For the residuals/predicted values plot we need to keep in mind that the NCO rates are theoretically non-negative. If we could precisely measure the NCO rates excluding other costs and regulatory considerations associated with charging off loans, they should always be between 0% and 100%. Therefore we can expect to see these distortions in the residuals.

The residuals for the Adaptive Lasso show some negative skew, where the RegARMA and benchmark residuals look normally distributed. Furthermore, Adaptive Lasso residuals are highly correlated. For the other models there appears to be a little auto-correlation in the residuals.

For the loan category Commercial & Industrial, we present the results of the Anderson-Darling Test, Engle's Arch Test, and Durbin-Watson test in Table 10.8.

We observe that the null hypothesis of no serial correlation in the residuals is rejected for all models at the 5% level. Normality could not be rejected for Adaptive Lasso residuals. Since all models suffer from auto-correlated errors, none of the models describe the complete average NCO rate process for Commercial & Industrial loans. This can be troublesome because in a stress test we assume that all conditions remain equal, but without a complete model specification we are not able to control

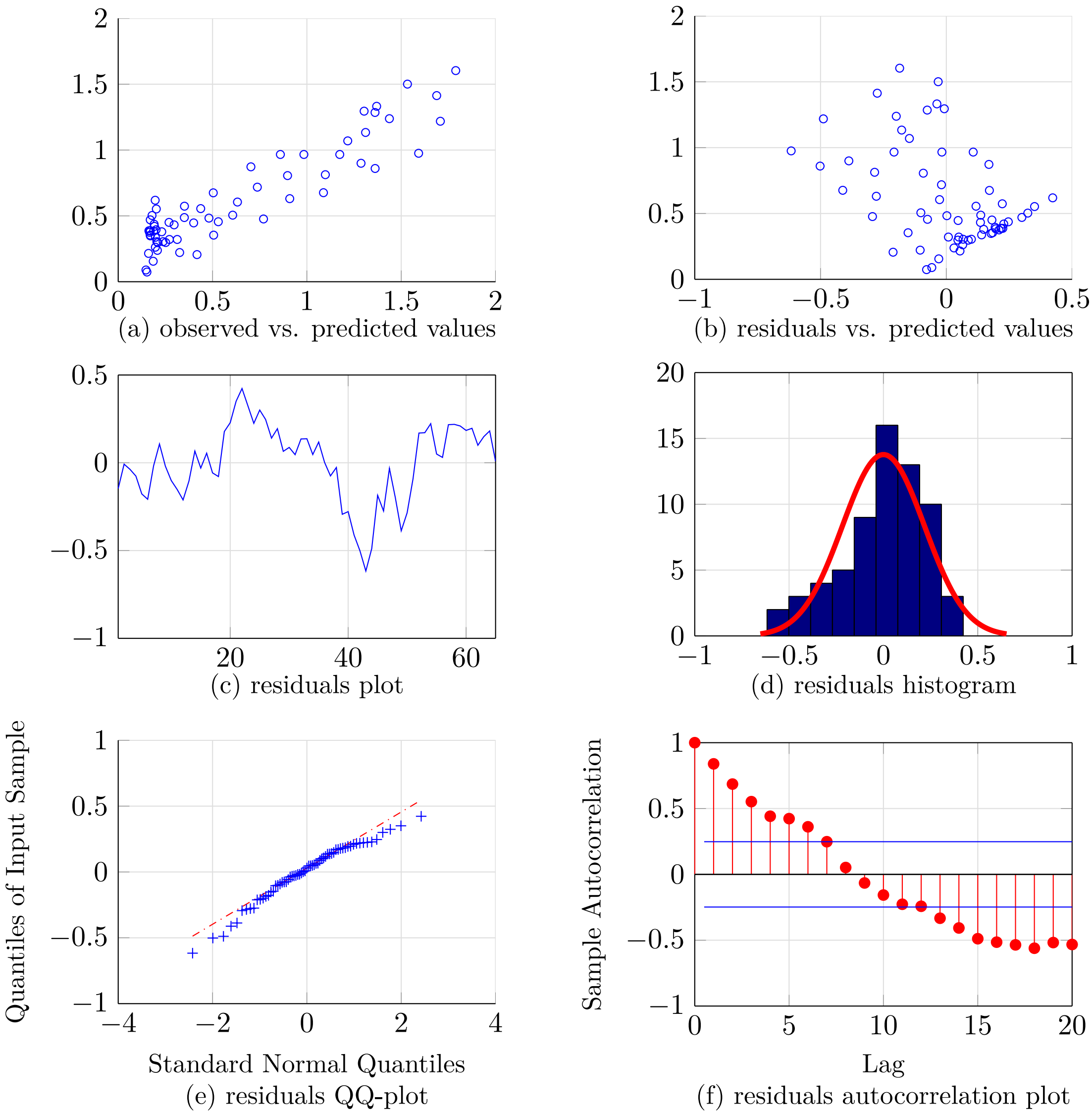
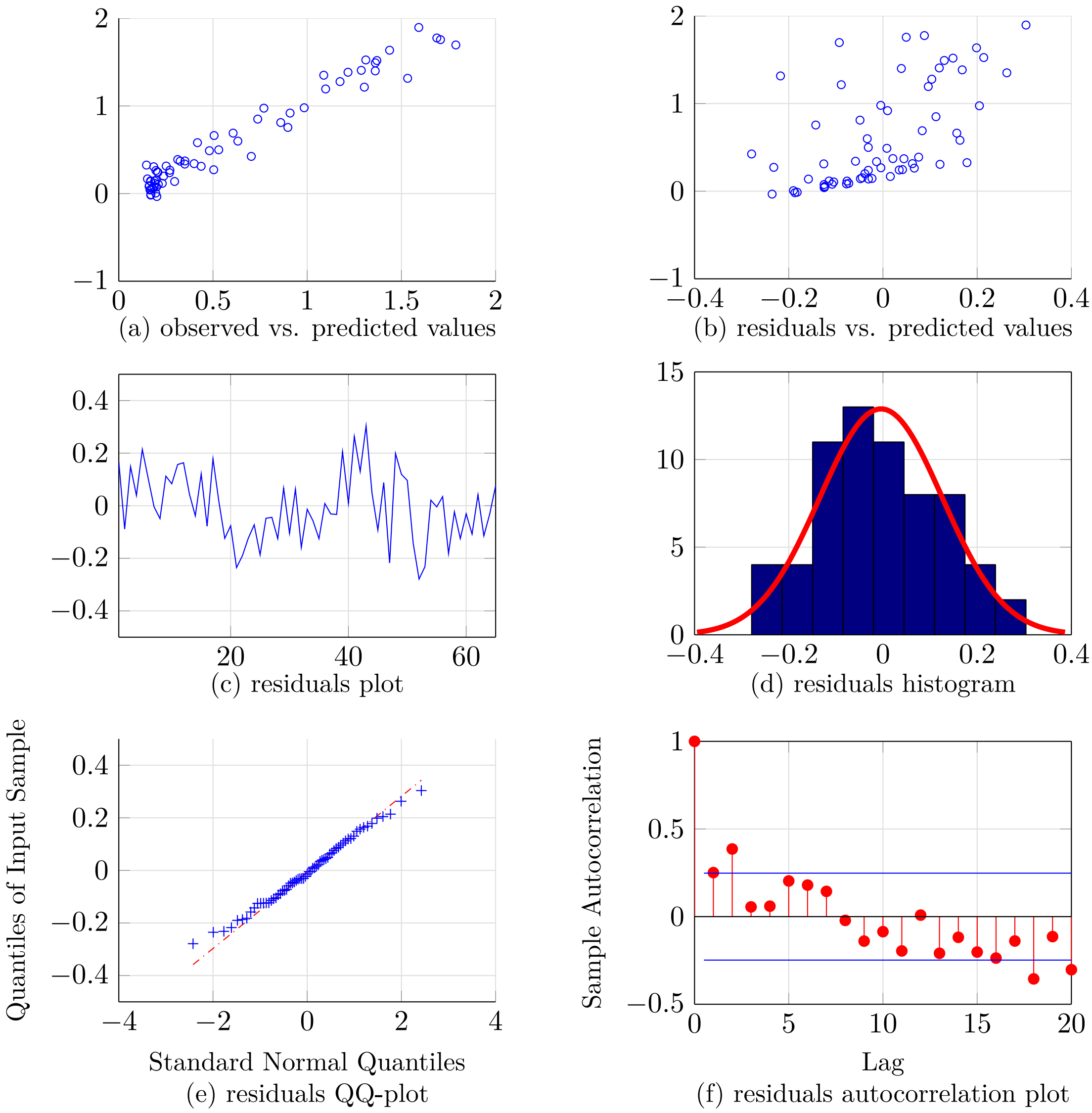
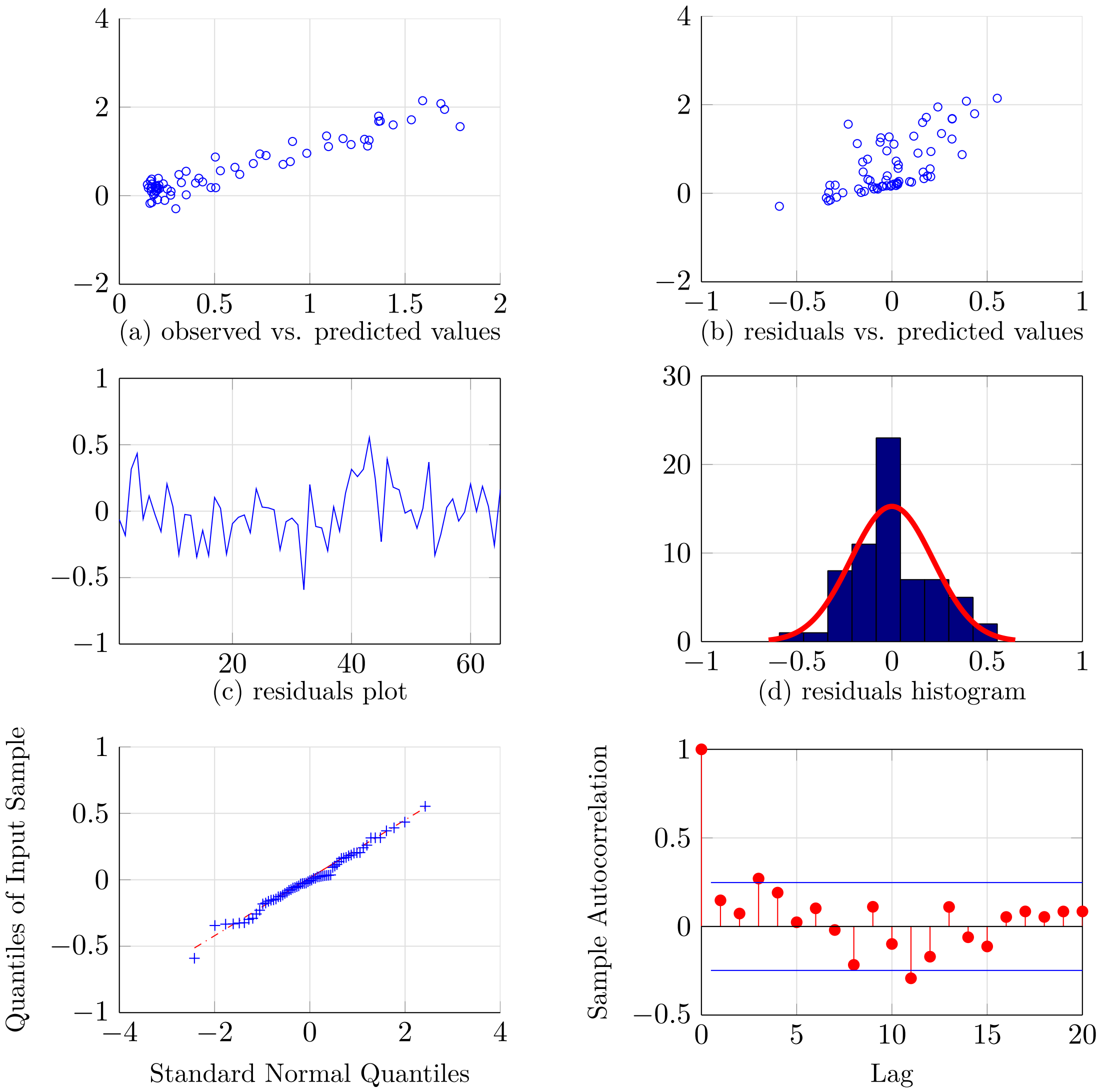


Figure 10.8: Residual Analysis for adaptive lasso on C&I NCOs



(a) observed vs. predicted values
(b) residuals vs. predicted values
(c) residuals plot
(d) residuals histogram
(e) residuals QQ-plot
(f) residuals autocorrelation plot

Figure 10.9: Residual Analysis for RegARMA model of C&I NCOs



(a) observed vs. predicted values
(b) residuals vs. predicted values
(c) residuals plot
(d) residuals histogram
(e) residuals QQ-plot
(f) residuals autocorrelation plot

Figure 10.10: Residual Analysis for Benchmark model of C&I NCOs

Table 10.8: C&I Residual Analysis

	Adaptive Lasso	RegArma	Benchmark
AD-test	0.0758	0.9652	0.5610
Arch-test	0.0000	0.9321	0.6439
DW-test	0.0000	0.0047	0.0462

[1] The p-values for conducting the Anderson-Darling (normality), Engle's Arch test (homoskedasticity), and Durbin-Watson (auto-correlation) test on the residuals of the fitted model on the first 65 time points.

for those conditions.

Although this indicates that the Adaptive Lasso top-down model for average NCO rates is not complete, this is not a surprising result. It is not clear whether a complex process, such as loss rates can even be condensed to a simple top-down model. This could also indicate that auto-regressive terms of the target variable need to be included in the model. The fact that the residuals of the RegARMA show little auto-correlation, suggest that a more advanced version of this model could give an accurate description of the NCO rate process.

In our discussion we shall elaborate further on the extensions of our method that are needed to make it a reliable top-down stress testing tool. For now, we focus mostly on the main purpose of the model, namely to establish the feasibility and accuracy of the prediction of NCOs in extreme but plausible scenarios for the macro-economy using data-driven methods.

10.2 Answers to the Research Questions

Adaptive Lasso selected and estimated parsimonious liner model descriptions of the data in training samples. It is able to reduce a design matrix consisting of over 10000 transformations of and interactions between macro-economic variables to a simple model specification containing only 5 predictors on average. The regularities in the training data are picked up efficiently and can be generalized on a test sample. It outperforms auto-regressive and simple linear models on nearly all our measures of prediction accuracy.

The models that we considered for the averaged NCO rates were all able to give accurate forecasts of the rates in the recession and recovery period between 2007-2014. Our results show that up to 70% of the variability in NCO rates can be explained by a top-down model. For stress identification, nearly 90% accuracy can be obtained. This suggests that it is feasible to use top-down models for stress testing.

The accuracy of the forecasts on a test sample of the macro-economy was very high, and for some loan categories even better than for the 9-quarters-ahead predictions. Since top-down models are very general and do not take into account very specific scenarios, they can capture the general impact of macro-economic developments on NCO rates. Our results strongly suggests that top-down models can be used to forecast NCO rates in stress test scenarios of more than five years. Especially Adaptive Lasso models were able to attain high accuracy.

For some averaged loan categories, the forecasts were not accurate when compared to the actual NCO rates. The main reason is that data-driven methods only pick up the regularities in the data on the training sample. If these regularities are not representative of the general NCO rate process, non-sensible predictions occur. Furthermore, when the scenario for a predictor variable that is relevant for forecasting rates on a certain loan category, is much more extreme than the data in the training data, forecasts are inaccurate. For reliable forecasts in stress testing, a minimum of at least one business cycle in the data is required. Since one economic downturn has specific characteristics, the general impact of a downturn can be captured better when two cycles are present in the data. The prediction accuracy of Adaptive Lasso on several loan categories suggests that the prediction error is smaller when two down-turns are in the training data. Depending on the economic history this means that approximately 20 years of data is necessary to estimate the impact of stress scenarios on all loan categories.

The use of a completely data-driven method to obtain top-down models for NCO rates thus sets limitations. Adaptive Lasso automatically summarizes the link between macro-economic developments and NCO rates in historic data. The resulting model can then be used to assess the impact of stress scenarios that are in the same range as the macro-economic variables in the training data. If the scenario is more extreme than the macro-economic developments that a model was trained on, the reliability of the forecasts cannot be guaranteed. Therefore a top-down model should only be used for a limited range of stress scenarios.

The auto-correlation in the residuals of the fitted Adaptive Lasso model suggest that it does not capture the complete NCO rate process. This suggest that an auto-regressive model can be better suited for the top-down modeling of NCO rates. However, the linear benchmark model showed the least auto-correlation in the residuals. Therefore, it could also be that the Lasso has auto-correlated residuals because it trades off between bias and variance when estimating the coefficients of a linear model.

Chapter 11

Individual Effects

In this chapter we present the results of the NCO rates model with individual effects. Recall the model including bank-specific effects,

$$\mathrm{NCO}_{i,j,t} = \alpha_j + c_j \left(M_t \beta + B_{j,t} \gamma + \epsilon_t \right), \tag{11.1}$$

The results of the Adaptive Lasso model with individual effects will be discussed in Section 11.1. Recall the research questions that we aim to answer with this model that were posed in Section 3.2.2,

i *Can bank-specific variables help explain variability in the credit losses between banks?*

ii *If we assume that $\gamma = 0$, can we then obtain accurate predictions for $NCO_{i,j,t+h}$? Or, is the inclusion of bank-specific effects necessary?*

The findings in Section 11.1 will be used to answer these questions in Section 11.2.

11.1 Model Results

We present the results for loan categories Commercial & Industrial and Construction & Land Development only. The first category is interesting because it has gone through two business cycles during the training data, and the latter is chosen because it is the only category for which one of the bank-specific CAMELS criteria in Table 5.2 was selected. All results were obtained by the Adaptive Lasso method for panel data, that we described in Section 8.2.

11.1.1 Selected Variables

For our 100 observations at 96 time points, Lasso only selected one of the CAMELS criteria from Table 5.2 on the loan category Construction & Land Development. The Adaptive Lasso models of NCO rates on the other loan categories did not contain bank-specific predictors. The selected models were very similar to those for the average NCO rates. The selected variables on the individual level are displayed in Table 11.1 below.

Table 11.1: Selected Variables on the Individual Level

C&I		CLD	
Variable	Sign	Variable	Sign
$\mathbb{1}_{\{\mathrm{GDP}_t<0\}}\mathrm{VIX}^3_{t-2}$	$+$	$\mathbb{1}_{\{\mathrm{HPI}_t<0\}}\overline{\mathrm{HPI}}^2_{t-4}$	$+$
$\overline{\mathrm{DUR}}_{t-1}$	$+$	$\mathbb{1}_{\{\mathrm{CPPI}_{t-1}<0\}}\overline{\mathrm{INF}}_{t-3}$	$+$
$\mathbb{1}_{\{\mathrm{VIX}_t>\mathrm{mean(VIX)}\}}\overline{\mathrm{HPI}}^2_{t-4}$	$+$	$\mathbb{1}_{\{\mathrm{GDP}_{t-1}<0\}}\overline{\mathrm{INF}}^2_{t-2}$	$+$
$\mathbb{1}_{\{\mathrm{GDP}_{t-1}<0\}}\mathrm{VIX}^2_{t-2}$	$+$	$\mathbb{1}_{\{\mathrm{UR}_t>\mathrm{NROU}_t\}}\mathrm{UR}_t\overline{\mathrm{DUR}}_{t-3}$	$+$
$\mathbb{1}_{\{\mathrm{DUR}_t<0\}}\overline{\mathrm{UR}}_{t-4}$	$-$	Net Interest Margin	$-$

A key observation is that indeed the model for Commercial & Industrial NCO rates is very similar to those for the average model displayed in Table 10.1. The first three selected variables are identical, save for a *lagged* averaged change in unemployment rate. The predictor $\mathbb{1}_{\{\mathrm{GDP}_{t-1}<0\}}\mathrm{VIX}^2_{t-2}$ is very similar to $\mathbb{1}_{\{\mathrm{GDP}_t<0\}}\mathrm{VIX}^3_{t-2}$. The last predictor indicates that when the unemployment rate is decreasing, then the change in NCO rate is negatively related to the average unemployment rate one year ago.

The selected variables for Construction & Land Development are conditional on a decrease in home prices, a decline in commercial property prices, a shrinking GDP, and an above average unemployment rate. Conditional on these events, the NCO rate is proportional to average home prices, inflation, and change in unemployment. From the CAMELS criteria, the Net Interest Margin was selected. Indicating that high margins on interest rates for individual banks are related to lower NCO rates.

Since the Net Interest Margin is highly correlated between banks, this could also suggest that financial indicators could be relevant in a top-down model of NCO rates. Therefore, it is prudent to include yields and interest rates in the initial design in the future. The fact that for all other categories none of the CAMELS criteria were selected by Adaptive Lasso indicates that these are not substantially relevant in explaining the variation in NCO rates between banks.

11.1.2 Prediction Accuracy

The predictions for NCO rates on Commercial & Industrial loans, based on the panel data model selected by Adaptive Lasso for the 'Big Four' commercial banks in the United States are shown in Figure 11.1.

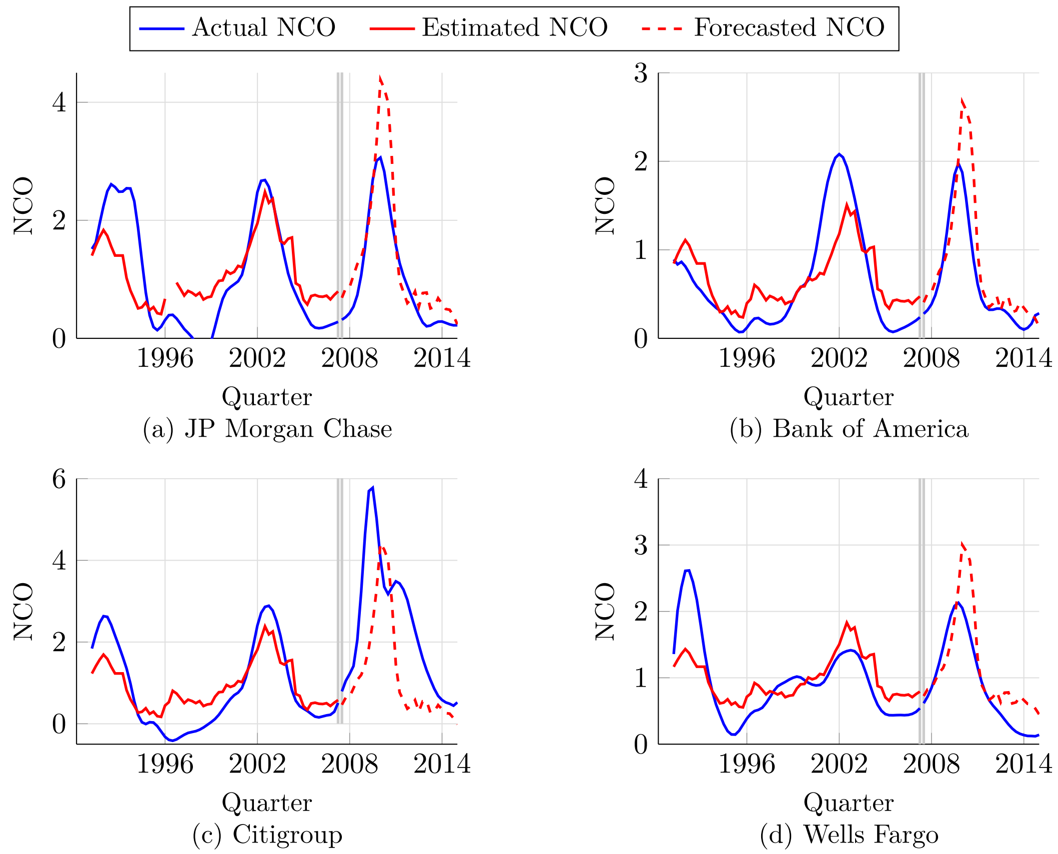
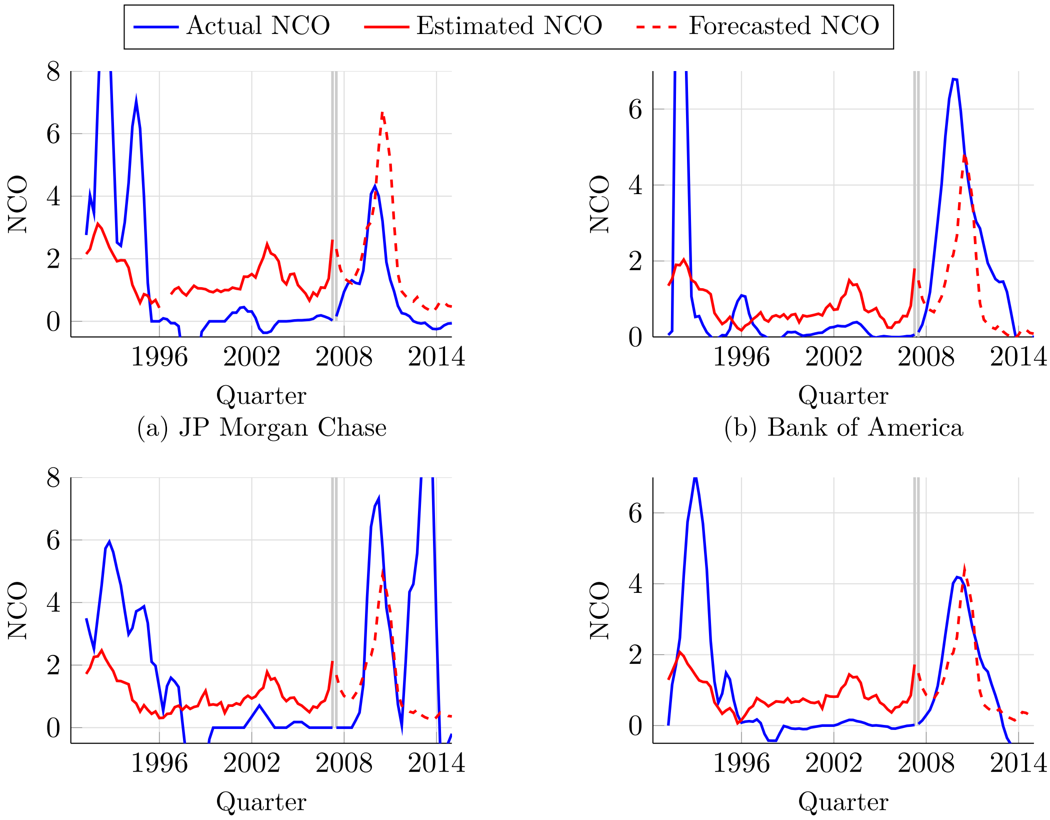


Figure 11.1: Estimated and Forecasted NCO rates on Commercial & Industrial Loans for Banks in the Test Panel.

We note that both the estimated and predicted NCO rates overestimate and underestimate the actual rates, depending on the bank. Since no bank-specific variables were selected and only the individual scaling factors α_j and c_j were used, this is to be expected. Since the dotted line seems to be able to capture the main movements in the NCO rates, the panel data model gives reasonably accurate predictions.

In Figure 11.2 the results for the prediction for Construction & Land Development NCO rates is displayed. The dotted line is not able to capture the bank-specific response of the net charge off rates, despite the inclusion of the bank-specific Net Interest Margin. Furthermore, the estimated NCO seems to be considerably biased. This may be caused by the fact that we optimized the tuning parameter λ in the

Adaptive Lasso estimation by minimizing the cross-validation error in the cross-sectional dimension. Our observations so far suggest that although the panel data model is able to predict the correct response of NCO rates generally, it is not very accurate and cannot capture bank-specific effects.



(a) JP Morgan Chase

(b) Bank of America

(c) Citigroup

(d) Wells Fargo

Figure 11.2: Estimated and predicted NCO rates on Construction & Land Development Loans for selected banks. The model was selected and estimated on the training data. The predictions are made a the vintage point of the beginning of the test set.

In Table 11.2 the prediction accuracy and strength of the model for NCO rates is displayed. The prediction accuracy differs greatly between the banks in our test panel. Since there are no bank-specific variables that influence the shape of the prediction, the accuracy depends greatly on how much the shape of the NCO rate resembles that of the average bank. Note that the predicted values explain much of the variation, as is indicated by a relatively high score of R^2_{efron} for C&I NCOs.

We conclude that the (untransformed) CAMELS criteria in the Adaptive Lasso model are not well-suited to model the cross-sectional dimension of NCO rates. Since the individual scaling factors α_j and c_j are important, it remains an open problem why these factors differ among banks, and what could cause them to change over time.

Table 11.2: Prediction Accuracy for C&I Individual on Test Set

	MSE		MAE		R^2_{efron}	
	C&I	CLD	C&I	CLD	C&I	CLD
JP Morgan Chase	0.1175	2.1349	0.4195	1.0320	0.6247	-0.1396
Bank of America	0.0351	3.6864	0.2611	1.4260	0.5168	0.5014
Citigroup	1.3306	13.258	1.2906	2.5512	0.4214	0.0700
Wells Fargo	0.0656	0.8314	0.3584	0.7967	0.5309	0.7033

[a] Coefficients are calculated based on normalized predictor variables, estimated on the training set.

11.1.3 Stress Identification

Besides the actual predicted values, we also measured the models ability to detect stress in the data. For stress threshold T_2, the results are presented in Table 11.3.

Table 11.3: Stress Identification for C&I Individual

	Spec.		Sens.		F1		Acc.		R^2_{count}	
	C&I	CLD[1]	C&I	CLD	C&I	CLD	C&I	CLD	C&I	CLD
JPMorgan Chase	0.96	0.88	0.83	1.00	0.83	0.77	0.94	0.90	0.67	0.40
Bank of America	0.96	1.00	0.83	0.53	0.83	0.70	0.94	0.77	0.67	0.53
Citigroup	1.00	0.94	0.35	0.47	0.52	0.61	0.65	0.71	0.35	0.40
Wells Fargo	0.96	1.0	0.71	0.73	0.77	0.84	0.85	0.90	0.57	0.73

[1] The threshold for Construction & Land Development was taken to be half the standard deviation from the mean. Because the peak in NCO rates for this loan category was much higher during the early nineties, relatively little stress is present if we take one standard deviation to identify stress.

We note that the panel data model is still able to detect the stress response of the NCO rates to macro-economic developments to a high degree. For *Citigroup* the results are somewhat disappointing, but if we look at figure 11.1c we can see that the shape of the NCO curve looks very dissimilar to those of the other banks. Since we were not able to capture bank-specifics apart from the individual scaling factors, it is to be expected that the accuracy for this time-series is not as good as for the other banks.

Also note that the evaluation of the model, where we compare the predicted values with observed values is not exactly comparable to the accuracy of a stress test. In that case, we assume that all circumstances besides the macro-economy remain equal. Therefore, the effects of idiosyncratic events are not predicted. The last peak in NCO rates of Construction & Land Development loans for *Citigroup* in Figure 11.2 could be caused by such an idiosyncratic event, or by the specifics of the Construction & Land Development portfolio. Therefore this does not necessarily mean that the top-down model failed to identify the expected NCO rate conditional on the macro-economy and *keeping all other relevant factors constant*.

11.2 Answers to the Research Questions

Adaptive Lasso did not select any of the CAMELS criteria in our panel data model for NCO rates. Only for NCOs on Construction & Land Development loans, was the Net Interest Margin selected. We interpret this result as an indication that financial indicators related to yields and interest rates should be included in the initial formulation in the future. We conclude that the inclusion of CAMELS criteria in the design matrix cannot explain the heterogeneity in NCO rates of different banks.

We found that individual scaling factors are able to explain much of the variability between NCO rates for different banks. The mean and range of the NCO rate data of banks in the United States differ greatly, but these difference appear to be somewhat constant over time. To give better forecasts of individual banks NCO rates on specific loan categories, it would be wise to develop methods to model these scaling factors. In order to gain a better understanding of the process it is interesting to discover the determinants of these factors. The fact that a model with individual scaling factors seems to work best for a top-down model is striking. Glasserman and Tangirala (2015) show that the *bottom-up* stress test results of the Federal Reserve's stress testing exercise are predictable. This is demonstrated by showing that the results for different banks have been linearly related over the last few years.

The results of this chapter suggest that NCO rates for specific loan categories can be forecasted accurately when individual scaling factors are taken into account. When we are able to separate the system-wide impact of the macro-economy on NCO rates from the idiosyncratic behavior of rates on specific loan portfolios, reliable top-down stress test results can be obtained under the same conditions as for the averaged case in the previous chapter.

Chapter 12

Individual Banks

In the previous two results chapters we established that Adaptive Lasso is able to accurately predict NCO rates. Individual effects on the NCO rates can best be described by the bank-specific scaling factors α_j and c_j. The remaining variation cannot be explained by the CAMELS criteria that we considered for the cross-sectional dimension. Recall the model with individual scaling factors,

$$\text{NCO}_{i,j,t} = \alpha_j + c_j \left(M_t \hat{\beta}^{\text{alasso}} + \epsilon_t \right). \tag{12.1}$$

The results are aggregated according to the method described in Section 8.3. The aim of this chapter is to explore whether the resulting top-down model for credit losses on the total loan portfolio for individual banks is feasible. Recall the research questions of Section 3.2.3.

i *What is the influence of the granularity of the NCO data on the accuracy of top-down models?*

ii *What prediction accuracy can be obtained by a top-down stress tests on the loan portfolio of an individual bank?*

iii *What stress horizons are feasible for top-down stress testing?*

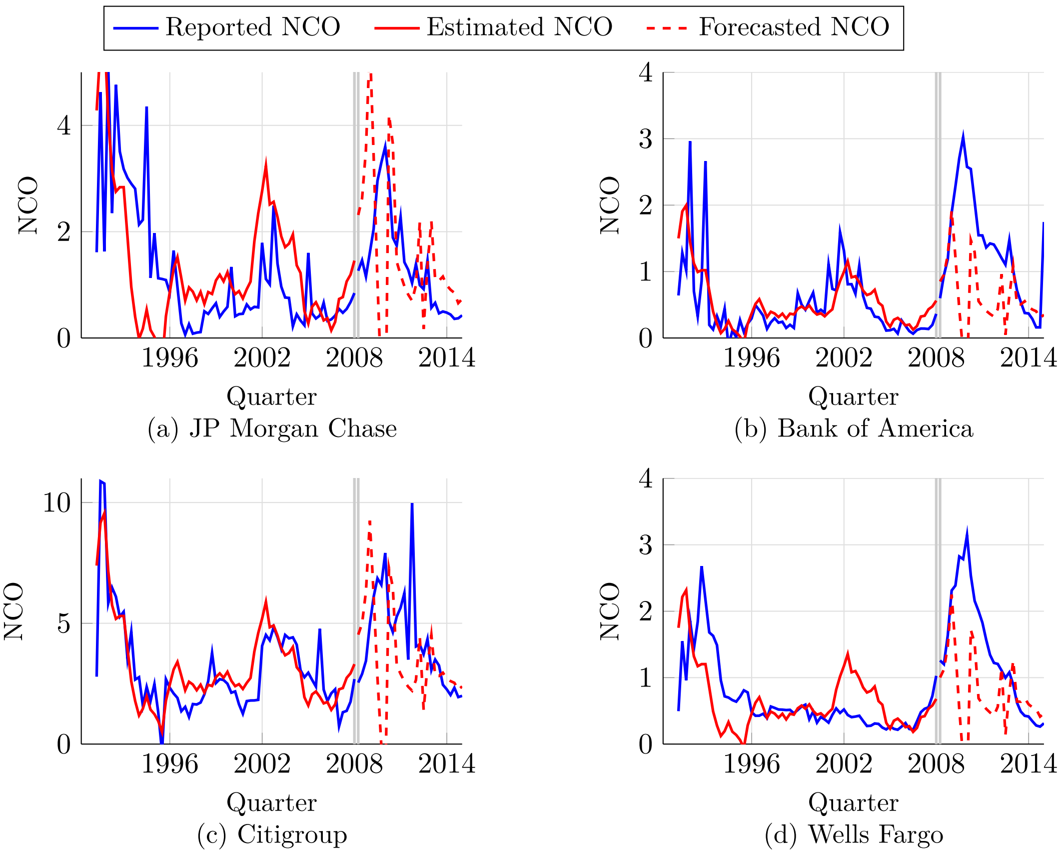
The results in Section 12.1 will be used to answer these research questions in Section 12.2.

12.1 Model Results

For the prediction of NCO rates for individual banks, an important consideration is the composition of the loan portfolio. Since credit losses tend to differ greatly between loan categories this signifies a large contribution towards improving the predictions on the individual level. In this chapter we compare the three aggregation levels that we discussed in Section 8.3 and defined in Section 6.1.1. Since the CAMELS criteria did not improve to the prediction accuracy of the top down model, we consider the average 9-quarter-ahead predictions here. These predictions are rescaled using the bank-specific scaling factors α_j and c_j, obtained by the method described in equation 8.26.

12.1.1 Prediction Accuracy

In Figure 12.1 the results for the top level of aggregation are shown, compared to the reported NCO values for the 'Big Four' banks in the United States: JP Morgan Chase, Bank of America, Citigroup, and Wells Fargo. The predictions are not consistent, look noisy, and do not resemble the predicted values well. The top level does not seem to be an appropriate choice for the prediction of NCO rates conditional on macro-economic scenarios.



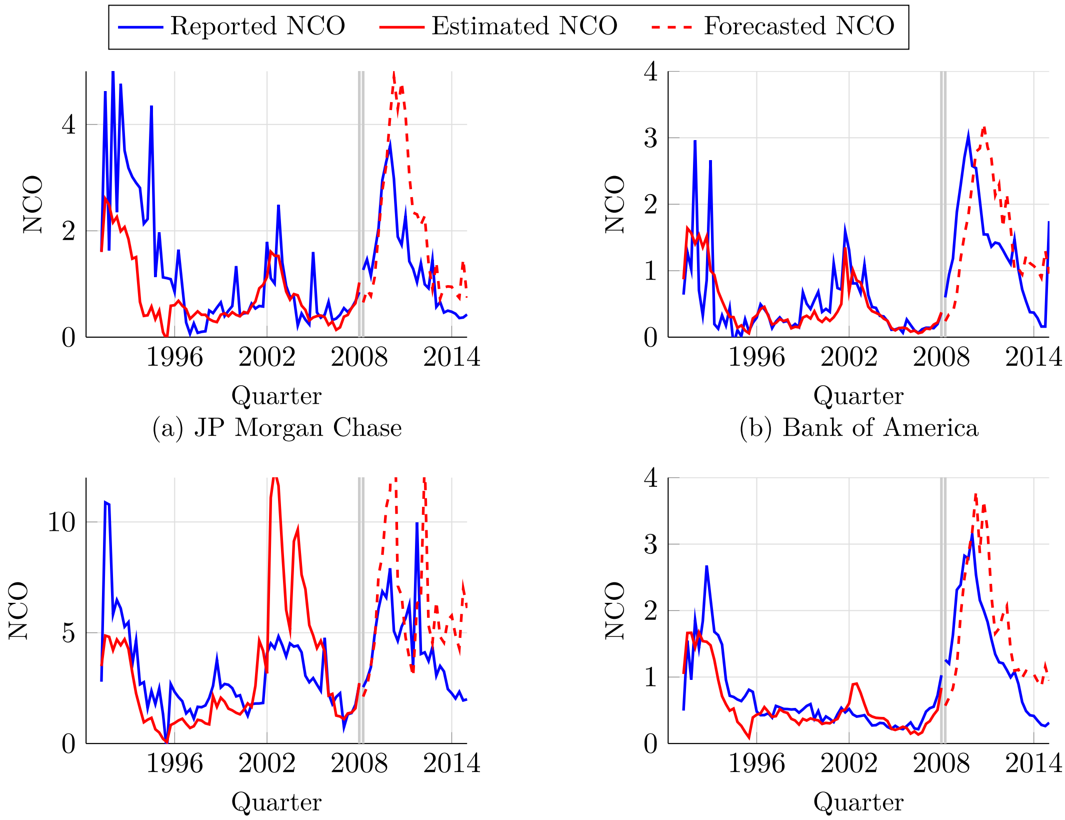
(a) JP Morgan Chase (b) Bank of America

(c) Citigroup (d) Wells Fargo

Figure 12.1: Aggregate Loan Portfolio Results for the top level.

Figure 12.2 shows the results for aggregation level A2. The dotted 9-quarter-ahead prediction resemble the reported values well, especially for JP Morgan Chase, Wells

Fargo, and Bank of America.



(a) JP Morgan Chase (b) Bank of America

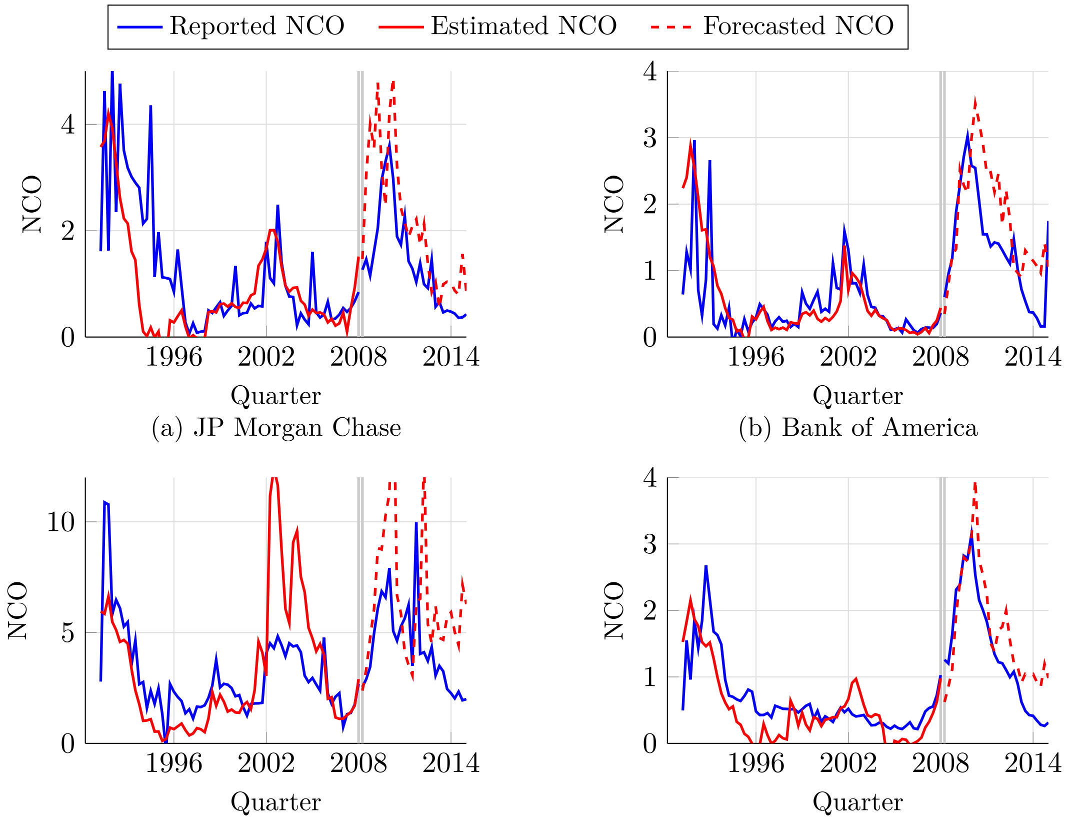
(c) Citigroup (d) Wells Fargo

Figure 12.2: Aggregate Loan Portfolio Results for bank panel at level A2.

Finally, the results for the bottom aggregation level A1 are displayed in Figure 12.3. Again, these seem to be very similar to the reported values. The results for Bank of America and Wells Fargo are even closer to the reported values than at the A2 level. For the reported NCO rates for Citigroup, the errors seem to be much larger than for the other banks in the panel.

The above comparison of the results for the four banks with the largest loan portfolio in the United States suggest that the level A1 and A2 are both suitable for a top-down model. It does seem however that the predictions are very sensitive to the estimated scaling factors for the individual banks. If these top-down models were to be used for a stress test, special attention needs to be paid to robust estimators for these bank-specific scaling factors.

The results in terms of MSE, MAE, and R^2_{efron} displayed in Table 12.1, make it easier to compare the three aggregation levels. The MSE has the highest value for level A3 for all banks in the test panel. The error is smallest for aggregation level



(a) JP Morgan Chase

(b) Bank of America

(c) Citigroup

(d) Wells Fargo

Figure 12.3: Aggregate Loan Portfolio Results for the bank level at the bottom level of aggregation.

A2 for JP Morgan Chase and Citigroup, where level A1 performs better on Bank of America and Wells Fargo. Since R^2_{efron} is negative for the banks for which A2 performs better, we conclude that the lowest level of aggregation is slightly better during the crisis period in the test sample.

Table 12.1: Prediction Accuracy on Aggregate Loan Portfolios

	MSE			MAE			R^2_{efron}		
	A1	A2	A3	A1	A2	A3	A1	A2	A3
JP Morgan	1.29	1.25	3.38	0.88	0.85	1.26	-0.50	-0.44	-2.91
Bank of America	0.59	0.75	1.40	0.68	0.78	0.81	0.57	0.44	-0.04
Citigroup	14.3	13.8	13.4	2.79	2.67	2.90	-1.63	-1.54	-1.47
Wells Fargo	0.31	0.55	1.38	0.46	0.64	0.75	0.75	0.56	-0.11

Note that the results for JP Morgan Chase in terms of prediction accuracy are caused by the overshooting of the actual NCOs before the peak at level A1 and after the peak at level A2. This is mainly caused by the rescaling that we used to obtain these results. Look at the high peaks at the beginning of the NCO series for JP Morgan Chase. These cause JP Morgan Chase to have relatively high scaling factors. For the stress test it was assumed that these remained the same as in the training data. Since we were not able to find variables to explain individual behavior we cannot explain why NCO rates for JP Morgan Chase were higher at the economic downturn in the early nineties than during the recession of 2007-2009.

12.1.2 Stress Identification

In terms of the correct identification of stress for the predictions we see a similar pattern, when we compare A1, A2, and A3 as for the prediction accuracy. The results in Table 12.2 indicate that A1 is best for Bank of America and Wells Fargo, and A2 is best for JP Morgan Chase and Wells Fargo.

Table 12.2: Stress Identification on Aggregate Loan Portfolios

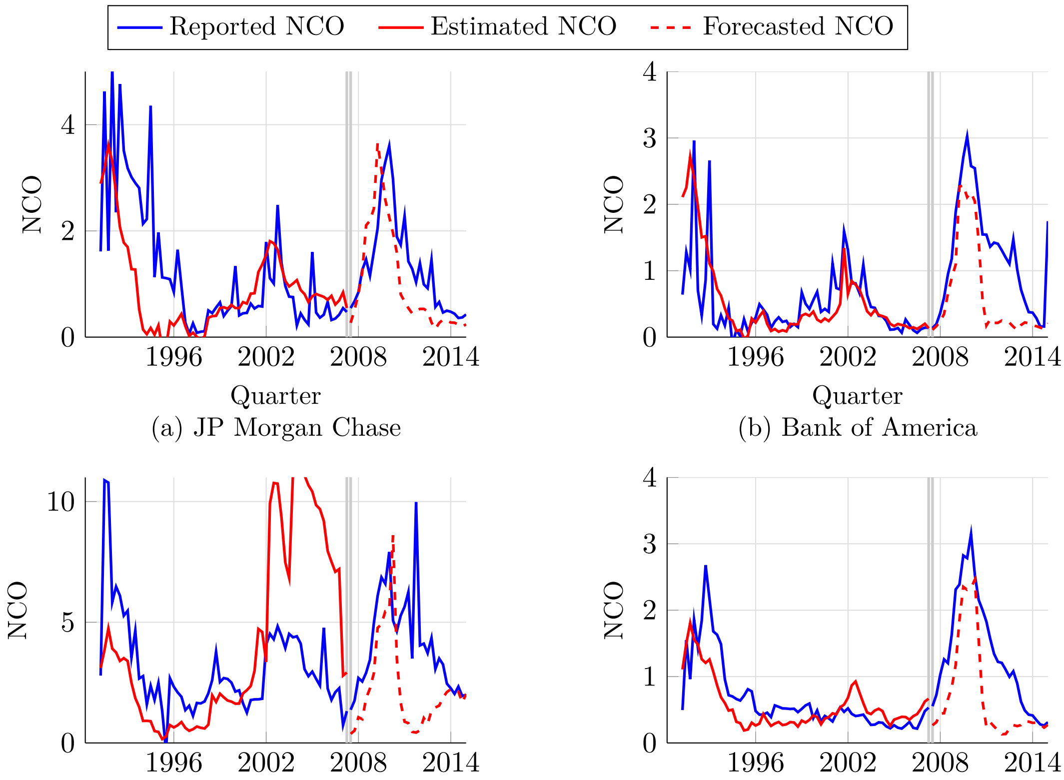
	Specificity			Sensitivity			F1		
	A1	A2	A3	A1	A2	A3	A1	A2	A3
JP Morgan	0.79	0.83	0.75	0.75	1.00	1.00	0.50	0.67	0.57
Bank of America	0.60	0.70	0.20	0.89	0.78	0.50	0.84	0.80	0.51
Citigroup	0.45	0.5	0.65	0.75	0.75	0.38	0.48	0.50	0.33
Wells Fargo	0.77	0.85	0.23	0.87	0.80	0.40	0.84	0.83	0.39

The accuracy in terms of sensitivity, specificity, and F1 is high. An average F1 score of 70% is attained for the lowest level of aggregation. This indicates that stress events can be captured with high accuracy.

12.1.3 Results for Long Stress Horizons

Since the prediction accuracy on the test data is most closely related to the goals and purposes of top-down stress testing, we show the results of the test panel on the test set for the 'Big Four'.

The results are slightly too optimistic, in the sense that higher NCO should be forecast, which is likely caused by the inability of the top-down model to capture the movements in the NCO rate for Closed-End Residential Real Estate loans. The same results can be obtained by using lagged values for Home Equity Lines of Credit and Closed-End Residential Real Estate NCO rates. We show the results on the test set for the bank in the test panel, for a model where these categories are incorporated as lagged variables.



(a) JP Morgan Chase (b) Bank of America

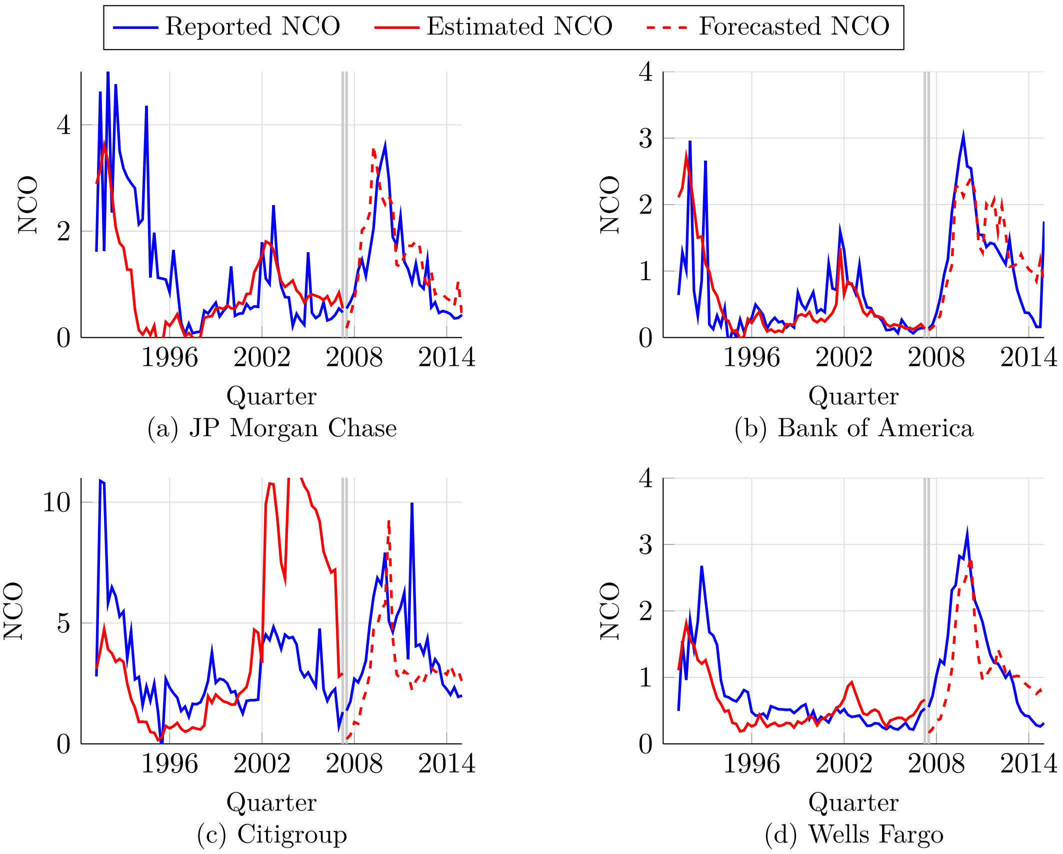
(c) Citigroup (d) Wells Fargo

Figure 12.4: Aggregate Loan Portfolio Results for the bank level at the bottom level of aggregation on the test set.

From our research we found that including a proxy for Home Equity Lines of Credit and Closed-End Residential Real Estate NCO rates, improves the prediction accuracy considerably. This suggest that a model is definitely needed that captures the response of these loan categories to macro-economic developments. If the data of the 2008/2009 financial would be included in the training set, it is likely that a new peak in losses for Closed-End Residential Real Estate loans would be predicted more accurately. This is also indicated by the selected variables on the entire dataset presented in Table 10.2 in Appendix E.

Finally, we are able to identify to which extend our implementation of a top-down model, tested on the 2008 financial crisis, resembles the realized stressed NCO rates. Since the predictions for the banks in the test panel are on a different scale, we calculated the *normalized* mean absolute prediction error (NMAE) and the normalize mean square prediction error (NMSE).

The final results for the 9-quarters-ahead predictions, and the forecasts for the test



(a) JP Morgan Chase
(b) Bank of America
(c) Citigroup
(d) Wells Fargo

Figure 12.5: Aggregate Loan Portfolio Results for the bank level at the bottom level of aggregation on the test set with a proxy for residential real estate NCO rates.

set with and without adjustment using lags values are displayed in Table 12.3. We observe that the results on the test set greatly improve when an adjustment is made to compensate for the error introduced by the lack of fit of *residential real estate* NCO rates. Furthermore, we see that the normalized mean absolute prediction error is 22% for the 9-quarters-ahead predictions and 18% on the test set. This supports the hypothesis that the stress horizon for top-down stress testing can be extended. Furthermore from the accuracy score in Table 12.3 we conclude that the top-down model that we developed accurately predicts stress for credit loss rates. Because our top-down method can predict NCO rates and identify stress accurately we conclude that top-down stress testing can be rewarding exercise and thus it is useful as a complement to bottom-up methods.

Remarkably, we can predict NCO rates with high accuracy for a stress horizon of nearly 8 years. A normalized mean square error of only 16% is obtained, using only data-driven methods to discover macro-economic relations. Furthermore, stress events are identified with more than 80% accuracy. These results are very promising

Table 12.3: Top Down Results

	9-quarters	Test	Adjusted Test
NMAE(%)	22	18	13
NMSE(%)	45	32	16
Accuracy(%)	73	77	81
Sensitivity(%)	84	40	66
Specificity(%)	66	99	90
F1(%)	72	57	72

for the further development of top-down methods.

12.2 Answers to the Research Questions

We compared the prediction accuracy for estimation of NCO rates on three different aggregation levels. Our results strongly indicate that the loan portfolio must be separated into a minimum of five distinct loan categories to capture the effect of macro-economic shocks. We find that distinguishing between the widely-used categories Residential Real Estate, Commercial Real Estate, C&I, Consumer, and Other Loans gives satisfactorily forecasts in a stress scenario. We note here, that further disaggregation seems to have a favorable effect on overall prediction accuracy. If we aggregate the portfolio into 12 categories, the error of our forecasts is only 16% and stress is predicted with more than 80% accuracy. Therefore, we conclude that the availability of granular data for estimation is key to reliable forecasts.

The stress horizon can be extended for more than five years, and high prediction accuracy can still be obtained for protracted horizons. In fact, the predictive strength does not even seem to be affected by lengthening the stress horizon. The scope of the training data remains an important consideration. Because of the extremity of the house price scenario during the most recent financial crisis, the house-price scenario in our test set was much more extreme than in the training data. This had a negative impact on our ability to capture the determinants of losses on real estate loans. Since our method is completely data-driven and these behaviors cannot be picked up from the training data, other inputs are required to give accurate forecasts for such scenarios. For the United States, the limitations set by the domain of the training data can be alleviated by using Adaptive Lasso on the complete data-set.

Part V

Conclusion & Discussion

Chapter 13

Conclusion

The focus of this report is on the development of methods to assess the impact of stressed macro-economic scenarios on credit losses in the loan portfolio of individual banks. Some tools that are currently available for such an exercise are expert opinions and data-intensive bottom-up stress tests. A top-down quantitative method that can be used to quickly predict the response of the credit losses on the loan portfolio of an individual bank, which can, for instance, be used to assess its financial position conditional on any scenario is a great complement to its risk framework.

Financial crises in the past follow similar patterns. This suggests that we can learn from past behavior to assess the propagation of sensible what-if scenarios to the loan portfolio of a bank. However, an often heard claim is that each crisis is different and unpredictable. The specific causes and propagation of a shock to the macro-economy or the financial system can typically not be known beforehand. Where bottom-up stress testing allows us to predict the response of credit losses to very specific shocks for stress horizons of approximately two years, top-down stress testing gives us the opportunity to give longer forecasts for more general scenarios. The macro-economy and its effect on loan losses are complex processes. At the start of this research project it was not clear whether a simple linear model would suffice to capture the link between macro-economic variables and credit loss rates.

The models that we developed in this report provide compelling evidence to the fact that a parsimonious linear top-down model can give accurate forecasts for the credit loss rates of banks during recession and recovery scenarios. Remarkably, the complex relation between the macro-economy and credit loss rates can be captured in linear models which typically have only $2 - 5$ predictors. The added bonus of such a simple model is that it is easy to interpret, and can thus contribute to a better understanding of macro-economic risks. Moreover, our simple models identify stress with an accuracy of 81%, and have normalized prediction errors of 16% for individual banks. A very favorable result is that our results suggests that these prediction accuracies can be obtained for stress horizons of five years or even longer. The implication is that a top-down model is able to forecast the general impact of stress scenarios for longterm planning.

We obtained these results on bank-data and macro-economic data from the United States. For a top-down model it is prudent to consider averaged credit loss rates over all banks. This way, we can adjust for bank-specific influences on credit loss rates and idiosyncratic events, by only considering system-wide changes in credit loss rates. It is important that the available data allows us to make this distinction.

To our knowledge there exists no publications of the effects of aggregation on prediction accuracy of credit loss rates. In this report we showed the influence of using granular loan data for estimation. That is, we compared the prediction errors and accuracy on three different levels of aggregation. The results show that it is required to consider at least five different loan categories. These can be described as Residential Real Estate, Commercial Real Estate, Commercial & Industrial, Consumer, and Other Loans. We have shown that higher prediction accuracy can be obtained by disaggregating further. This shows that the loan composition of an individual bank is an important determinant of its credit loss rates.

The major contribution of this thesis to the development of stress testing methods, is the introduction of a completely data-driven approach. We show that Adaptive Lasso can be used to discover a parsimonious top-down model from a set containing thousands of possible model specifications. This is particularly useful when theoretical models are not available, as is the case with top-down stress testing. We considered many potentially relevant transformations of macro-economic variables and their interactions in an initial formulation of the design matrix. We applied Adaptive Lasso as a learning algorithm to select the most relevant predictors.

In this report we showed that the Adaptive Lasso procedure always gives sparse solutions, is approximately unbiased, and has attractive convergence properties. Importantly, it can distinguish between many different model specifications and gives unique solutions, even when the number of potential predictors is much larger than the number of observations. The predictive strength of Adaptive Lasso models was much higher than that of benchmark models selected by classical methods. As far as we know, this is the first application of Adaptive Lasso to bank data.

Our results indicate that Adaptive Lasso automatically picks up regularities in the training data and searches for those variables that describe the behavior of credit loss rates best. This results in a parsimonious description of the relation between the macro-economy and credit loss rates. The fact that the method is completely data-driven means that it only considers those events that are present in the training data. In a sense, it gives us a summary of past events in terms of a linear model specification. Important effects of the macro-economy can therefore not always be identified on a small data-sample. The results presented in this thesis suggest that the link between macro-economic development and system-wide credit loss rates generally follows a regular pattern, but at least one, and preferably two, economic downturns are required for reliable forecasting. This suggests that roughly 20 years of quarterly data is necessary for the development of top-down stress testing methods. Another option is to limit the scope of the stress scenarios used on a top-down model.

Another consideration during the research project was the inclusion of bank-specific variables to explain heterogeneity in NCO rates for different banks. We included bank-specific criteria and transformations of macro-economic predictors in a panel data model and used Adaptive Lasso to select the substantially relevant predictors. Interestingly, the Net Interest Margin was selected as an indicator in one instance. This suggests that interest rates and yields could be relevant predictors for credit loss rates. To our knowledge none of the existing top-down stress testing model for credit losses includes such predictors in its specification. For the remaining part, none of the bank-specific criteria were selected. The macro-economic predictors that were selected were very similar to those selected for the average loan case, which further supports the robustness of Adaptive Lasso. Based on our findings we are of the opinion that the adaptation of a panel data model is not worth the effort. But note that our search for suitable bank-specific variables far from exhaustive.

The mean and range of credit loss rates differed among banks and captured the variability between different banks well. This suggests that individual scaling factors are an important part of a top-down model. It would be interesting to characterize the determinants of the differences in mean and variance in credit loss rates for different banks. This allows us to understand the underlying causes. Lastly, the Adaptive Lasso model was compared to a model with the same predictors but with auto-regressive error terms. The residuals of the fitted Adaptive Lasso model were auto-correlated and those of the auto-regressive model were not. On the other hand, the linear Adaptive Lasso model had higher prediction accuracy on the test sample.

Recommendations The models that we discovered and tested, are based on data from the United States. The method that we used rendered simple top-down models, which can be interpreted with ease, and give accurate longterm forecasts in stress scenarios. This makes it attractive to pursue the adaptation of such methods for banks in Europe and more specifically, in the Netherlands.

A closer look at our experiences gives insight in how such a simple top-down model can be constructed. Our method is completely data-driven and obtains a parsimonious top-down model by learning from past data. Therefore, it can easily be adapted to learn a model on a different economic system or for differently specified stress scenarios. The main issue is that a substantial amount of data is needed to train a model which can give reliable longterm forecasts.

Our findings suggest that granular data on credit losses is needed. To that end, we require historical credit loss rates for the loan categories Residential Real Estate, Commercial Real Estate, Commercial & Industrial, Consumer, and Other. In order to capture the link between the macro-economy and credit losses on these loan categories in a data-driven manner, we need to be able to identify regular behaviors. To that end, quarterly data is required for a period of roughly 20 years, or it should contain at least one complete business cycle. Because top-down stress testing requires us to determine the systemic impact of stress scenarios, the credit loss rate data need to be representative of the entire banking system. In conclusion, granular quarterly data considering system-wide credit loss rates are needed for a period of 20 years.

Understandably, the above data requirements can be challenging for the near future. Therefore, we recommend to consider using other inputs for the selection and estimation of a top-down model. In the final chapter of this report we make some suggestions as to how this can be done.

Chapter 14

Discussion

The results in the previous part of the thesis indicate that a top-down stress testing tool is feasible as long as the amount of estimation data is sufficient and the stress scenarios that serve as input for the prediction of rates is not too extreme (as compared to the macro-economic developments in the estimation data). In that case, only five parsimonious linear top-down models for an aggregated loan portfolio are sufficient to give accurate forecasts for credit loss rates in the stress scenario. Furthermore, efficient methods using Adaptive Lasso are available to flexibly select an appropriate model, which can be evaluated by experts. This is a promising result.

In this chapter we raise four distinct issues. First, we take the opportunity to give an overview of possible improvements to our method that could lead to even better results. We proceed by taking a critical look a the assumptions that we explicitly and implicitly made in the modeling process. Then we raise some important questions about the available evaluation methods for top-down macro-economic models. We conclude with a discussion about the concerns associated with the domain of the training data.

14.1 Improvements

Here we identify some possible improvements to our method, that gives the opportunity to use it as a top-down stress testing tool. Our suggestions concern the range of the credit loss rates, the inclusion of financial macro-economic indicators, the construction of confidence bands, and further research towards explaining bank heterogeneity in NCO rates. But we start off, by considering the quality of the NCO data.

Quality of the Data Since we are modeling NCO *rates*, which are the annual percentage losses for each dollar that is loaned out in a certain loan category, the range of the response variables in our model should theoretically be between $0-100\%$. But due to the limitations set by the reporting method, negative rates do occur in NCO data. Furthermore, unknown costs and decision-making processes are involved with writing off debt in the banking book, as we discussed in Chapter 6. Since these factors were unknown to us, we treated these effects as measurement errors, and used a seasonal filter to remove them. The seasonal averaging approach that we employed, is perhaps too crude. It uses 4 extra data-points and could filter out relevant information. With a better understanding of the source of the non-random signal due to regulatory and seasonal effects, a more advanced filtering method can be employed.

Another consideration is that we used average NCO rates to capture the system-wide impact of macro-economic shocks to credit loss rates. Although this approach removes fluctuations due to circumstances unrelated to macro-economic developments, it requires a lot of data. Therefore, it could be worthwhile to consider other methods to obtain observations of system-wide credit loss rates.

Range of NCOs The values of the *actual* NCO rate takes on values within a certain range. When the range of the output of a model is limited to this range, prediction accuracy can be greatly improved. The models that we developed can give negative predictions for NCO rates. Furthermore, the residuals of our fitted models showed some distortion for observed values close to zero. It would be prudent to consider a transformation suitable for the modeling of percentage data.

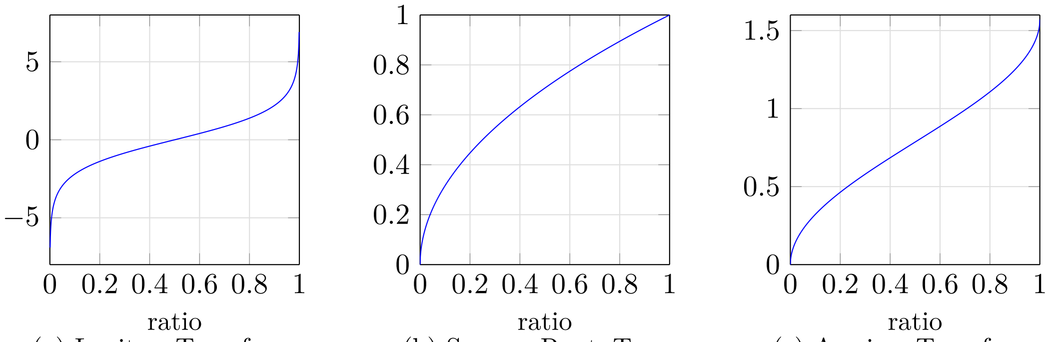
A widely-used transformation that maps percentage data to the real line is the logit transformation, given by

$$\log\left(\frac{p}{1-p}\right), \tag{14.1}$$

where $0 \leq p \leq 1$.In the case that the percent data fall in the range 0 to 10% it is often recommended to use a square root transformation on the data instead. Another frequently used method is to apply an arcsin transformation. The effects of these transformations are displayed in Figure 14.1.

Our data mostly lies in the range $0-2.5\%$. We found that both transformation have a distorting effect on the shape of the time-series. Therefore, we decided to model the untransformed data. Further study is needed to find a suitable transformation, which ensures positivity of the model's predictions. This might also have a positive influence on the distribution of the residuals.

The Macro-Economic Effect From our panel data model using bank-specific effects we found some evidence that suggests that financial indicators such as interest rates and bond yields should be included in the initial formulation. That way,



(a) Logit Transformation
(b) Square Root Transformation
(c) Arcsin Transformation

Figure 14.1: Commonly used transformations that ensure that the response variable is between 0 and 1.

Adaptive Lasso can select these variables if they are substantially relevant predictors of NCO rates.

An important consideration is the down-scaling of the initial formulation. Although our results indicate that Adaptive Lasso is capable of distinguishing between substantially relevant and unimportant variables, it would be prudent to only include variables that are deemed economically significant. Macro-economic indicators, transformations, or interactions that are very unlikely related to credit loss rates can be removed from the initial formulation based on expert judgment and economic theory. Furthermore, the linear models that are simultaneously selected and estimated by Adaptive Lasso should be evaluated thoroughly from an economic point of view.

The Individual Effect We used two approaches to include individual effects that explain the variability in NCO rates between banks. First of all, we introduced individual scaling factors, to scale the predictions of our models based on the historic mean and variance of NCO rates for individual banks. These scaling factors were calculated for each loan category. An important consideration is the estimation method of the bank-specific scaling factors which influence the level and range of NCO rates for individual banks. Determining the underlying principle of these scaling factors, could give further insight in the risk drivers of credit losses. Then, forecasts in macro-economic stress scenarios can be scaled differently when characteristics of the loan portfolio are adjusted.

As a second approach to capturing individual effects, we considered bank-specific CAMELS criteria in a linear panel data model. Our results showed that these did not explain the cross-sectional variation in NCO rates, and we concluded that a panel data model is not worth the effort. But our search for explanatory variables was far from exhaustive, and it could be pursued further. If bank-specific variables can explain the cross-sectional variation, this could be a step towards a complete top-down model that describes all of the NCO rate behavior. Finally, the panel data model could be improved by considering interactions between bank-specific variables and the macro-economy. Note that a panel data model requires a lot of bank-specific data. Since our results show that the individual scaling factors explain much of the heterogeneity, it is much more important to investigates these factors further.

Confidence Bands For the practical use of a stress test it is necessary to give some estimate of the reliability of the prediction by providing confidence bands. However, for the models that we considered for this research project it is not easy to determine accurate values for the confidence bands. Firstly, the necessary conditions on the NCO rate process for inference using t-tests, F-tests, etc. are not satisfied. And secondly, the degrees of freedom in an automatically selected model based on cross-validation or BIC values cannot be determined.

For the Adaptive Lasso model in this report, there exists several suggestions in literature to calculate confidence bands. These methods are based on the convergence rate of a procedure that possesses the oracle properties. However, these measures can only be calculated in the $n < p$ case. Therefore, to calculate confidence bands the initial formulation should be reduced significantly or large amounts of data need to be made available.

Another option to give a measure for reliability of the predictions is to use the prediction error on a test sample. Unfortunately, this requires more data since we need to split our data in a training and a test sample. As we have seen on our dataset at least 20 years of data is already needed for the training sample. A solution might be the use of cross-validation techniques.

14.2 Assumptions

To verify the possibility of using top-down models for stress testing credit loss rates and in order to establish the extend of the accuracy of such methods, we developed a top-down stress testing method. If such a top-down model were to be used for actual stress testing, the assumptions that we made along the way need to be addressed. These assumptions were explicitly and implicitly stated in this report and can be summarized as follows.

Model Specification An important aspect of the model is its functional form. For the Adaptive Lasso model in this report we assumed a linear model specification.

We also considered a model with auto-regressive error terms and predictor variables selected by Adaptive Lasso. The prediction error of this auto-regressive model, was slightly higher than that of the linear Adaptive Lasso model. Because the in-sample fit of the auto-regressive model was much better, it is worth considering.

In the auto-regressive model case, Adaptive Lasso methods can be used to simultaneously select and estimate both the macro-economic predictors and the lags of the model. The issue with including lags and moving averages of error terms in a specification for Adaptive Lasso, is that the minimization problem is no longer convex. Hence finding the solution is significantly more computationally challenging. Wu and Wang (2012) put forward an approximation scheme to efficiently select and estimate such a model with Adaptive Lasso in [33].

Correlations For simplicity we assumed that the distribution of NCO rates and the individual scaling factors remains the same over time. Based on the information in our data-set this assumption is reasonable, but this gives no guarantees for modeling in the future.

We cannot be sure that the measurement error in the data that we used is small. A better understanding of how the NCO data is gathered, could alleviate this issue. Neither are we certain whether there is correlation between NCO rates on different loan categories besides a common dependency on the macro-economy. For a panel data model we should consider whether there exists correlation between NCO rates for individual banks, besides their common dependency on macro-economic developments.

Adaptive Lasso In order to discover regularities in our data, we assumed that the true model is sparse and used Adaptive Lasso. If the true model is not sparse, no model can describe and predict the behavior of NCO rates conditional on macro-economic paths.

Another considerations is the fact that we used $\gamma = 1$ for the weighting of Adaptive Lasso. For future applications, it is wise to optimize this parameter of Adaptive Lasso. For this thesis we used cross-validation techniques on the training sample to select the tuning parameter λ with the smallest prediction error. But other and perhaps better methods are available, which can be explored.

In the application of Adaptive Lasso to the linear panel data model, we used that $\lambda_1 = \lambda_2$ for simplicity. In the future the optimal values for λ_1 and λ_2 should be selected. Recall that for cross-validation the data is split into folds to determine the prediction error on each of these folds. When cross-validation is used on panel data to select tuning parameters, the creation of folds becomes more challenging. More research is needed to establish best practices.

Macro-Economic Modeling One of the main issues in macro-economic modeling, and also in our top-down stress testing model, is that the assumptions for regression are not satisfied for the estimation data. Most importantly, the observations are not independent and identically distributed (i.i.d.) samples. Moreover, it is not likely that the measurement error of all variables is negligible. Or that the processes involved are stationary and non-changing over time. This is one of the realities of macro-economic modeling.

We do not expect that the NCO rates are completely determined by the values of the macro-economic stress variables in our initial formulation. It is likely that other (macro-economic) factors and idiosyncratic events have an impact on the NCO rates for individual banks. These are not included in our model.

14.3 Evaluation of Top-Down Models

The extend of top-down stress testing was investigated by assessing the accuracy of the Adaptive Lasso method on the financial crisis, recession and the following recovery of the last eight years. We are aware that our results for prediction accuracy and stress identification were obtained on only one test sample. Therefore, the reliability of the prediction errors is uncertain. The major issue in macro-economic modeling is however, that the possibilities for the evaluation of models are limited. We have a few suggestions how more reliable estimates of prediction accuracy can be obtained.

First, the response of NCO rates in a top-down model can be tested by using historical scenarios. A second option is to compare the forecasts of a top-down model with those of bottom-up methods. Lastly, it can be tested on the stress test scenarios that the Federal Reserve prescribes annually or on self-made stress scenarios. The predicted response can then, for instance, be compared to the opinion of experts.

Although methods for evaluation are scarce, a more thorough approach to the testing of top-down models is feasible. In our opinion it is necessary to acquire reliable estimates of the predictive strength of a top-down model in order to use it as a top-down stress testing tool.

14.4 The Domain of the Scenario

For NCO rates on loan categories which are explained by variables whose value during the 2007-2009 crisis scenario was close to values attained in the training sample, the predicted rates are adequately estimated. In the results part of this report we observed that the forecasts for the loan categories Closed-End Residential Real Estate, Farmland, Home Equity Lines of Credit, and Agriculture are particularly inaccurate. An obvious explanation for this is that the macro-economic developments of the 2007-2009 crisis had not been observed before. In the data between 1990-2006 the house prices had nearly always increased, and therefore the effect of a decline in house prices could not be captured accurately by any of our data-driven models.

A data-driven top-down model cannot magically predict the future and can only use past data to estimate the relation between macro-economic changes and credit losses. It is therefore important to realize what an extreme scenario is for a specific loan category and how this impacts the reliability of the stress test results. In other words, the model only shows how the NCO are related in the domain of the training data, and does not necessarily give good predictions outside of this domain. It is therefore useful to be aware of whether you are extrapolating beyond the boundaries of the models too much. And it is prudent to quantify the domain of the macro-economic variables in the training data. Then it can be determined whether the range of the variables used in the stress scenario are in the domain of the training data. When the scenario is too extreme, the forecasts are unreliable.

Because the data requirements for the development of data-driven top-down stress testing methods are substantial, we took the opportunity to explore some options to extend the domain of the model in other ways. In the next and final chapter of this report we discuss some methods that can be adopted that use alternative inputs for the construction of a top-down model.

Chapter 15

Extending the Domain

For this research project, the aim was to assess the feasibility of designing an adequate top-down stress testing method, to determine the prerequisites for such an exercise, and to develop methods to implement it. A question that evolved naturally in the course of the research project, was how to deal with small data-sets (i.e. data-sets on which it is not feasible to estimate a complete top-down stress testing model.). To that end, we searched for methods that can be used to generalize models estimated on data of other economic systems. Our attention was directed to the use of other inputs that can be used for the model selection and estimation procedure.

15.1 Alternative Input Data

A viable option is to extend the domain of a top-down model by quantifying the input of expert judgments. Expert opinions can be used to give an idea of what happens in the extremal points outside the scope of the training data. This allows us to interpolate predictions between those of the expert inputs at the extreme and those of the data-driven model. In this case the condition that a top-down model should be data-driven needs to be relaxed.

Another possibility is to use data for the macro-economy on a smaller scale. For the United States, we have seen that there is no observation of declining house prices prior to the 2007 subprime-mortgage crisis in our data-set. This has consequences for the ability of our model to learn the correct response to declining house prices, which becomes clear when we look at the prediction accuracy for the loan categories Residential Real Estate and Home Equity Lines of Credit (HELOC). But in some states or cities, such a situation might have occurred previously. Based on the response of credit losses in these regions, we might be able to estimate a model on this partial data set. The important question here is, how this model can be generalized to the entire nation-wide bank population.

Another option is to use historic data. It is widely acknowledged that the older the data is, the less representative of the current situation it is. Perhaps transfer learning techniques may also be useful to learn from events that occurred in a time that the market conditions were drastically different from the current situation. If we are able to retain the relevant information from observations of different but similar systems over time, location or scope, we can extend the domain of a top-down stress testing model.

An interesting topic for further research is whether the restrictions on top-down stress testing that are due to constraints imposed by the domain of the training data can be alleviated by using data from other (related) systems. For instance, can the knowledge that we derive from the markets in the United States, be used to better understand the developments in European markets? There is little doubt, that typically, experts indeed use knowledge of processes that they learned to understand from foreign markets. It would be interesting to determine whether this process can be replicated quantitatively. A new approach would be to borrow techniques developed in the field of *transfer* learning.

15.2 Transfer Learning

This promising area of research focuses on the development of learning algorithms that use labeled or unlabeled data from a system with a different distribution to improve the learning curve of the program in another system. In the situation that one has two datasets with sampled from different distribution but closely related, and the data that one wants to predict is scarce, it would be convenient to employ the data in the secondary dataset to improve the prediction accuracy. A classical example of transfer learning is that when you have learned how to play the piano, it will be much easier to learn to play the organ than if you had not obtained that skill. The question is whether the same reasoning can be applied to machine learning algorithms. As part of our discussion, we give a short introduction to some techniques that may be used to transfer knowledge that was learned on a different system, to enhance learning on the economic system.

Although transfer learning currently is an increasingly popular subject in machine-learning contexts, the literature on the application of transfer learning to economic models is scarce. In [24] data from heterogeneous sources is scaled in order to make predictions in the target set based on data in a source domain. The economic example of such a situation that is provided is the prediction of down-town housing prices, using labeled data on suburban house prices. In [29] the call report data from the U.S. is used in a transfer learning algorithm to predict bank failures in Australia.

In the context of supervised or regression learning, transfer learning strives to solve the following problem. Consider two datasets containing labeled data, that were sampled from different distributions. Let the target dataset be denoted by $(\mathcal{X}^T, \mathcal{Y}^T)$, and the source dataset by $(\mathcal{X}^S, \mathcal{Y}^S)$, where $\mathcal{Y}$ are the labels and $\mathcal{X}$ the features of the model. For our case, we assume that both the domain and the learning tasks differ. Hence if the target data is distributed according to $p^T(x, y)$ and the source data as $p^S(x, y)$, it holds that $p^T(x, y) \neq p^S(x, y)$. This type of transfer learning is commonly referred to as *inductive* transfer learning.

One option is to transfer knowledge by boosting for regression transfer. The main idea is to learn hypotheses for a model on the source data. Each hypothesis can be seen as an expert. There are several schemes to combine these experts to construct a model for the target data. Such schemes weight the hypothesis in such a way that the prediction error on the labeled target data is minimized. This process is referred to as *boosting for regression transfer* [23].

To give an impression of how transfer learning might be implemented we summarize an introduction to another method that transfers knowledge for a regression learning task. We consider an importance weighted inductive transfer learning method that can be found in [45]. Here it is key to assess which data-points contribute positively to the prediction accuracy on the target data and to discard data points that have a negative influence.

The solution that importance weighted inductive transfer learning [45] proposes to this problem is to estimate a measure of similarity between the source and target distributions. We shall see that this comes down to estimating a so-called importance weight function $w(x, y) = \frac{p^T(x,y)}{p^S(x,y)}$, which can be used to reweight the data in the source domain,

$$p^T(x, y) = w(x, y)p^S(x, y). \tag{15.1}$$

Note that for predictions $\hat{y}$, it holds that

$$\hat{y} = \arg\max_y \left(p(y|x)p(x)\right). \tag{15.2}$$

Since we have two distributions in the inductive transfer learning setting we can obtain two predictions $\hat{y}_S^T$ and $\hat{y}_S^T$ based on the target and source data, respectively. By equation 15.2,

$$\begin{aligned} y_T^T &= \arg\max_y (p^T(y|x^T)p^T(x^T)), \\ y_S^T &= \arg\max_y (p^S(y|x^T)p^S(x^T)). \end{aligned}$$

These prediction can differ, due to the difference of $p^T(x, y)$ and $p^S(x, y)$. The idea is to reweight the distribution p^S with the weight function $w(x, y)$. This can be rewritten like this,

$$
\begin{aligned}
y^T &= \arg\max_y \left(w(y, x^T) p^S(y|x^T) p^S(x^T) \right) \\
&= \arg\max_y \left(\frac{p^T(y, x^T)}{p^S(y, X^T)} p^S(y|x^T) p^S(x^T) \right) \\
&= \arg\max_y \left(p^T(y|x^T) p^T(x^T) \frac{p^S(y|x^T) p^S(x^T)}{p^S(y, x^T)} \right) \\
&= \arg\max_y \left(p^T(y|x^T) p^T(x^T) \frac{p^S(y|x^T) p^S(x^T)}{p^S(y|x^T) p^S(x^T)} \right) \\
&= \arg\max_y \left(p^T(y|x^T) p^T(x^T) \right).
\end{aligned}
$$

From the above derivation it can be seen that this gives an unbiased estimator for y^T. Since the amount of target data is limited, the estimation of $p^T(y, x^T)$ is difficult, but if there is a large quantity of source data it is feasible to estimate $p^S(y, x^T)$. With the use of the weight function, the estimation procedure comes down to estimating the weight function and $p^S(y, x^T)$. Note, that the weight function can be interpreted as a measure of similarity between the distributions. For an example of how $w(x, y)$ can be estimated, we refer to [45].

In our research project we have shown that a top-down stress test is feasible for stress scenarios that are within the scope of the domain of the training data. The available macro-economic and bank-specific data for the task of estimating a top-down stress testing model on European or Dutch data is typically scarce. However, the amount of data that is available from different but likely related economic systems is much larger. There are several directions to go from here, one could consider using bank- and macro economic data from geographically different regions in the world, or employing data from the 1970s and 1980s, which are typically too differently distributed from the current data. Finally, it could be helpful to use detailed data on a level where the impact of macro-economic developments is distinguishable.

Although some of the examples of the results that the transfer learning approach yield are promising, it remains unknown whether this type of knowledge transfer can work for credit losses of individual banks conditional on macro-economic stress variables.

Part VI

Appendices

Appendix A

Time Series

In this appendix we give a short refresher of general theory about, and notation that is used for, time-series. In particular, we introduce the convenient back-shift operator that we used in Chapter 8. Furthermore we give definitions of the ARMA, RegARMA, and the white noise processes that were mentioned. Finally, we discuss some key aspects of seasonality and decomposition of time-series.

A time-series is a sequence $\ldots, Y_{-2}, Y_{-1}, Y_0, Y_1, Y_2, \ldots$ of random variables. The series Y_t is indexed by equally spaced time points t, where the implied ordering of the variables is essential. Time-series are also commonly referred to as (stochastic) processes[1].

The study of time series involves both probabilistic and statistic methods. The probabilistic part is the study of the probability distribution of sets of variables Y_t that will typically be dependent. The statistical part is to study the probability distribution of the time-series given observations $Y_1, \ldots, Y_n$. The resulting model can be used to understand the underlying process or to make future predictions $\hat{Y}_{n+1}, \hat{Y}_{n+2}, \ldots \hat{Y}_{n+h}$ [40].

For more convenient notation, we introduce the back-shift operator B. We let B signify the transformation $B(y_t) = y_{t-1}$, where $B^n(y_t) = y_{t-n}$. Another useful transformation is the *polynomial lag operator* which performs the following operation

$$\begin{aligned} \phi(B)y_t &\equiv (1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)y_t \\ &= y_t - \phi_1 B(y_t) - \ldots - \phi_p B^p(y_t). \end{aligned}$$

In order to estimate probability distributions of a process, we need to assume that the time-series satisfy some structure. A commonly used structure is stationarity.

[1] Actually, the theory for discrete-time stochastic processes can be extended to continuous time. For top down modeling for stress testing, we do not consider this option, since only discrete time data is available.

When the underlying structure of a stochastic process does not change over time, it is called stationary. This is formulated more precise in Definition A.1 below. Another important concept in time-series analysis is white noise, defined in Definition A.2.

Definition A.1. (Stationarity) Let F_Y be the cumulative distribution function of the process Y_t. Let $t_1, \ldots, t_k$ represent k equally spaced time-points. We say that the time series Y_t is *strictly* stationary when

$$F_Y(y_{t_1}, \ldots y_{t_k}) = F_Y(y_{t_1+h}, \ldots y_{t_k+h}),$$

for all $h \in \mathbb{N}$. We say that the time series Y_t is *weakly* stationary when

i $\mathbb{E}[y_t] = \mu$, $\forall t$,

ii $\mathrm{Var}(y_t) = \sigma_x^2$, $\forall t$,

iii $\mathrm{cov}(y_t, y_{t-s}) = \sigma_s$, $\forall t, s$.

Definition A.2. (White noise) We say that a series ϵ_t is white noise if,

i $\mathbb{E}[\epsilon_t] = 0$,

ii $\mathrm{Var}(\epsilon_t) = \sigma^2$, $\forall t$,

iii $\mathrm{cov}(\epsilon_t, \epsilon_{t-s}) = 0$, $\forall s \neq 0$.

We use the above definition to introduce auto-regressive (AR) and moving-average (MA) processes. Let ϵ_t be a white noise process. An auto-regressive process of order p is denoted by AR(p) and is defined by

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \epsilon_t,$$

where ϵ is a white noise error, c is a constant, and ϕ_i are the parameters of the model.

A moving-average process Z_t of order q is denoted by MA(q) and is defined as

$$Z_t = \mu + \sum_{i=0}^{q} \theta_i \epsilon_{t-i},$$

where μ is the mean of the series, θ_i are the parameters fo the model, and ϵ_{t-i} are white noise errors. Note that where shocks ϵ in an AR-model influences the values of X infinitely into the future, for the MA(q) model the shock to Z is only propagated for q time periods.

A process that has both AR-terms and MA-terms is usually referred to as an ARMA(p, q) process. In Definition A.3 we give the precise statement.

Definition A.3. (ARMA(p, q) process) Let ϵ_t be a white noise series. We say that Y_t is an ARMA(p, q) process when,

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t.$$

In terms of polynomial backshift operators the above equation can conveniently be rewritten as,

$$\phi(B)Y_t = c + \theta(B)\epsilon_t. \tag{A.1}$$

In some settings there are exogenous variables that influence the time-series process. Such processes can be modeled by using an auto-regressive moving-average model with exogenous variables, which is denoted by ARMAX(p, q).

Definition A.4. (ARMAX(p, q) process) Let Y_t is an ARMAX(p, q)-process with exogenous predictors X_t if,

$$Y_t = \alpha + X_t\beta + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t. \tag{A.2}$$

In terms of the polynomial backshift operators, ARMAX(p, q) the model can be written in the more compact form,

$$\phi(B)Y_t = X_t\beta + \theta(B)\epsilon_t.$$

This can be rewritten to,

$$Y_t = \alpha + \frac{\beta}{\phi(B)} X_t + \frac{\theta(B)}{\phi(B)} \epsilon_t.$$

Note that the AR-coefficients are both related to the covariates and the error term. The coefficients therefore become difficult to interpret from an economic point of view. Exogenous variables that influence the time-series process can also be included in a linear model where the errors are an ARMA(p, q) process, we refer to such a model as a RegARMA(p,q) model.

Definition A.5. (regARMA(p,q) process) A regARMA(p, q) process Y_t is of the form,

$$\begin{aligned} Y_t &= \alpha + \beta X_t + \eta_t \\ \eta_t &= \sum_{i=1}^{p} \phi_i \eta_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t. \end{aligned}$$

In terms of polynomial backshift operators this becomes,

$$Y_t = \beta X_t + \frac{\theta(B)}{\phi(B)}\epsilon_t. \tag{A.3}$$

Notice how the coefficients for the covariates and the error terms are now separated. This greatly simplifies the interpretation of the estimated model parameters. The time-series models that we have discussed so far, decompose the time-series based on predictability. It turns out that a decomposition into a deterministic and purely indeterministic part is possible for *any* weakly stationary process.

Theorem A.6 (Wold's Theorem)**.** *Let X_t be a weakly stationary process with $\mathbb{E}X_t = 0$, then*

$$X_t = \sum_{j=0}^{\infty} \theta_j \epsilon_{t-j} + \eta_t, \tag{A.4}$$

where

i $\epsilon_t = X_t - \mathbb{E}[X_t|X_{t-1}, X_{t-2}, \ldots]$ *is a white noise process,*

ii η_t *is a deterministic process,*

iii $\sum_{j=0}^{\infty} |\epsilon_j| < \infty$,

iv *all the roots of* $\theta(L)$ *are on the unit circle,* θ_j*, and* ϵ_s *are unique.*

Proof. The theorem can be found in most books on time-series analysis. We recommend [57]. □

According to Wold's decomposition theorem all weakly stationary processes can be decomposed into a deterministic and a random part, justifying the model structures discussed above. But not all processes are stationary.

The most common types of non-stationarity in time-series analysis occur due to the presence of a linear trend or seasonality. A time-series may exhibit seasonality in the sense that it has a regular pattern which is repeated over s time periods. For example, for quarterly data it typically holds that $s = 4$ and for monthly data $s = 12$. Trend and seasonality can be removed from a time series by applying (seasonal) differencing, $X'_t = \nabla_s X_t = X_t - X_{t-s}$ for time-series with seasonality s. To detrend a time-series we calculate $X'_t = \nabla X_t = X_t - X_{t-1}$. Time-series are often stationary after (seasonal) differencing, if this is not the case then more differencing is generally required.

We can describe patterns in a time series using a simple method called classical decomposition, which is an important technique for time-series analysis and seasonal adjustment in particular. The main idea is that one assumes that the time-series can be decomposed in the following four parts [5]:

i The trend T_t describes long term movements of the mean.

ii The seasonality S_t relates to cyclical movements corresponding to calendar dates.

iii Cycles C_t describe other cyclical fluctuations.

iv The residuals ϵ_t contain other random movements.

It is then assumed that these four elements can be combined either additively or multiplicatively,

$$X_t = T_t + S_t + C_t + \epsilon_t \qquad X_t = T_t \cdot S_t \cdot C_t \cdot \epsilon_t. \tag{A.5}$$

Time-series decomposition is a frequently used component in texts concerning forecasting of business and (macro) economic data. Researchers are often interested in separating the infamous business cycles and growth cycles that are visible in macroeconomic data from seasonal and random components. This allows them to study the effects and interactions of business and growth cycles. Under the assumption that classical decomposition can be applied to (macro-economic) processes, these can be obtained by using filters or smoothing[2].

[2]For a further discussion of classical decomposition for macro-economic time-series we recommend [10].

Appendix B

Linear Regression

In this appendix we derive the normal equations for least squares regression, show why the number of predictors should be smaller than the number of observations in linear regression, discuss the assumptions of ordinary least squares regression and provide an introduction to the modeling of panel data in Section B.1. In the subsequent section we focus on forecasting with linear regression models. In particular we will discuss some measures for model evaluation, and we stress the importance of finding a model that can be generalized beyond the training set of data. Two measures that strive to perform this task are the AIC and BIC measures, which were also introduced in Section B.2.

B.1 Ordinary Least Squares Regression

An important tool in time-series modeling in the presence of exogenous predictor variables is linear regression. Some processes are determined by the paths of exogenous variables or predictors. Let X_t be a vector of observations of predictor variables and Y_t the value of the response variable at time t. Assume that we have n observations of response-covariate pairs. A linear model for Y_t is of the form,

$$Y_t = \alpha + X_t\beta + \epsilon_t, \, \forall \, t = 1, \ldots, n.$$

The α term can be included in the design matrix X, and then the vector β is extended with β_0. Therefore the constant term α is often dropped, which we shall do here as well. A popular method to estimate the coefficients β for such a linear model is Ordinary Least Squares (OLS) regression, which minimizes the sum of squared residuals in a linear model of the form

$$Y = X\beta + \epsilon, \tag{B.1}$$

where X is a $(n \times p)$-matrix and Y is a vector of length n. The squared residuals of the linear regression are given by,

$$\begin{aligned}
\epsilon^\top \epsilon &= (y - X\hat{\beta})^T (y - X\hat{\beta}) \\
&= y^\top y - y^\top (X\hat{\beta}) - (X\hat{\beta})^\top y + (X\hat{\beta})^\top (X\hat{\beta})) \\
&= y^\top y - 2(X\hat{\beta})^\top y + (X\hat{\beta})^\top (X\hat{\beta})) \\
&= y^\top y - 2(X\hat{\beta})^\top y + \hat{\beta}^\top X^\top X\hat{\beta}.
\end{aligned}$$

The minimum of these squared residuals is determined by deriving the first and second order derivative of the above expansion. The derivatives are given by

$$\frac{\partial}{\partial \beta}[\epsilon^\top \epsilon] = -2X^\top y + 2X^\top X\hat{\beta} \tag{B.2}$$

$$\frac{\partial^2}{\partial \beta^2}[\epsilon^\top \epsilon] = 2X^\top X. \tag{B.3}$$

In order to determine the minimum we set the first derivative to zero, and we require that the second derivative in equation B.3 is positive definite. From linear algebra we know that $X^\top X$ is positive definite if and only if X has full column rank. If this is the case, then by setting equation B.2 to zero, we find that the OLS regression has the closed-form solution

$$\hat{\beta} = (X^T X)^{-1} X^\top y. \tag{B.4}$$

When $p > n$ we know that X cannot have full column rank. Hence $X^\top X$ is not positive definite and therefore a unique solution to the least squares problem does not exist. In fact, $\hat{\beta}$ is not *identifiable* when X does not have full column rank. This can easily be shown. If we assume that X_t does not have full column rank, then without loss of generality we have that $X_1 = \alpha X_2$. In this case no distinction exists between,

$$y = \beta_0 + \sum_{k=1}^{p} \beta_k X_k + \epsilon, \tag{B.5}$$

and

$$y = \beta_0 + (\beta_1 \alpha + \beta_2) X_2 + \sum_{k=3}^{p} \beta_k X_k + \epsilon. \tag{B.6}$$

The Ordinary Least Squares regression is based on several assumptions on the underlying data $(y_1, X_1), \ldots, (y_n, X_n)$. Under some extra conditions the OLS method can be shown to be unbiased and have minimum variance among all estimators. The typical assumptions in ordinary least squares regression are:

i The model is linear in the parameters.

ii The residuals are statistically independent, $\mathrm{cov}(\epsilon_i, \epsilon_j) = 0$, $\forall\, i \neq j$.

iii The independent variables are not too strongly collinear.

iv The measurement error is negligible.

v The expected value of the residuals is zero, $\mathbb{E}[\epsilon] = 0$.

vi The residuals have homoskedastic variance.

vii The residuals are normally distributed, i.e. $\epsilon \sim \mathcal{N}(0, \sigma_2 \mathbb{I})$.

The first four assumptions are necessary for any estimation procedure of the regression parameters. The fifth assumption is a sufficient condition for the resulting OLS-estimator $\hat{\beta}^{\mathrm{OLS}}$ to be unbiased. An estimator depends on the values of the observations, if the expected value of the estimator is the actual parameter, we say that it is unbiased,

$$
\begin{aligned}
\mathbb{E}\hat{\beta}^{\mathrm{OLS}} &= \mathbb{E}\left[(X^\top X)^{-1} X^\top y\right] \\
&= (X^\top X)^{-1} X^\top \mathbb{E}[X\beta + \epsilon] \\
&= (X^\top X)^{-1} X^\top X\beta \\
&= \beta.
\end{aligned}
$$

Under the sixth assumption, the famous Gauss-Markov theorem states that the OLS estimator has the minimal variance of all linear unbiased estimators. The normality of the residuals allows us to use t-test and F-test for inference. We discuss some methods to test whether assumptions five to seven are satisfied. A thorough discussion of the validity, advantages, and disadvantages of these tests is beyond the scope of this research project, instead we give a quick overview of the tests that were used in this report.

A popular method to test the null hypothesis that residuals are from a normal distribution is the Anderson-Darling test. This test measures the distance between the empirical distribution function and the hypothesized distribution function. The test statistic can be calculated by,

$$
A = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{F(x)(1 - F(x))} dF(x). \tag{B.7}
$$

The Anderson-Darling test-statistic measures the squared distance between the empirical and hypothesized distribution but puts more weight on the tails of the distribution because it is divided by $F(x)(1 - F(x))$. The value can be compared to a critical value. When it is exceeded, the null hypothesis that the data belongs to the distribution F, is rejected.

A frequently used test statistic for the null hypothesis that the residuals of a linear regression do not have serial correlation is the Durbin-Watson test,

$$d = \frac{\sum_{i=2}^{n}(e_t - e_{t-1})^2}{\sum_{i=2}^{n} e_t^2}. \tag{B.8}$$

If the residuals are not correlated, then the value of the Durbin-Watson test statistic is close to 2. The null hypothesis of no serial correlation can be tested by comparing to lower and upper critical values, which can be calculated from the design matrix X.

We can test for auto-regressive conditional heteroskedasticity (ARCH) in the residual time series by using Engle's test. To determine whether a residual time-series ϵ_t contains ARCH-effects, the main idea is that a p-th order ARCH-model for ϵ_t can be written as

$$\sigma_t^2 = \mathbb{E}[\epsilon_t^2|\epsilon_{t-1}, \ldots \epsilon_{t-p}] = \gamma_0 + \gamma_1\epsilon_{t-1} + \ldots + \gamma_p\epsilon_{t-p}. \tag{B.9}$$

The null hypothesis of no ARCH-effects is tested by performing the auxiliary regression,

$$\hat{\epsilon}_t^2 = \gamma_0 + \gamma_1\hat{\epsilon}_{t-1}^2, \ldots, \gamma_p\hat{\epsilon}_{t-p}^2. \tag{B.10}$$

and calculating the following test statistic,

$$R^2 = 1 - \frac{\sum_{t=1}^{n_T}(\hat{\epsilon}_i - \epsilon_i)^2}{\sum_{t=1}^{n_T}(\bar{\epsilon}_i - \epsilon_i)^2}. \tag{B.11}$$

The test statistic is then given by, $n \cdot R^2$, which can be shown to follow a χ_p^2 distribution under the null hypothesis. Then if the test statistic exceeds critical values the null hypothesis of no ARCH-effects can be rejected.

A panel data model combines time-series with a cross-sectional dimension. Longitudinal time-series analysis allows us to simultaneously estimate time-series on data for different individuals, institutions, etc. We let the individuals be indexed by $j = 1, \ldots, n_B$, and time be indexed by $t = 1, \ldots, n_T$. For a *random* effects model, the analysis is based on a linear regression like this,

$$Y_{j,t} = X_{j,t}\beta + Z_{j,t}\gamma + u_j + \epsilon_{j,t}. \tag{B.12}$$

In the above equation the individual deterministic effect, or heterogeneity, is given by $Z_{j,t}\gamma$. This term represents individual-specific variables that are assumed to be either constant or evolving over time. The term u_j represents and individual random effect.

The *fixed* effects panel data model is given by,

$$Y_{j,t} = \alpha_j + X_{j,t}\beta + Z_{j,t}\gamma + \epsilon_{j,t}, \tag{B.13}$$

To find a solution (α, β, γ) to the panel data model, given response-covariate pairs $(Y_{j,t}, X_{j,t}, Z_{j,t})$, we need to solve,

$$\arg\min_{\beta,\gamma} \sum_{i=1}^{n_B} \sum_{t=1}^{n_T} \left(Y_t - \sum_{j=1}^{p} X_{t,j}\beta - Z_{t,j}\gamma \right)^2 . \tag{B.14}$$

We note that the same methods that are used for normal linear regression can be employed to solve the panel data least squares minimization problem in equation B.14.

B.2 Forecasting with Linear Regression

It is well known that forecasting errors increase typically increase when more variables are included in the model. When the complexity of a model increases, its bias decreases, but in the mean while the variance in the estimation of its coefficients increases. The aim in forecasting with linear regression is to optimize this trade-off between bias and variance in such a way that the prediction error is minimal.

In order to gain a better understanding of the so-called bias-variance trade-off, we again start with the classical setting. Assume that we have n observations of response-covariate pairs, $(y_1, x_1), \ldots, (y_n, x_n)$, where $y_i \in \mathbb{R}$ and $x_i \in \mathbb{R}^p \, \forall i = 1, \ldots, n$. We consider a linear model of the form

$$Y = X\beta + \epsilon, \tag{B.15}$$

where $Y = (y_1, \ldots, y_n)^\top$, $X = (x_1, \ldots, x_n)$ and $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^\top$ is a vector of i.i.d. random variables with mean 0 and constant variance σ^2. In this typical regression problem we want to find an estimator $\hat{\beta}$ for the true model β^0, which minimizes the prediction error, which can be accomplished by minimizing the Mean Square Error (MSE).

Let $f(\hat{\beta})$ denote the predictions given by the estimated model, and $f(\beta^0)$ those of the true model. Then the mean square prediction error can be written as,

$$
\begin{aligned}
\mathrm{MSE}(\hat{\beta}) &= \mathbb{E}\left[\left(f(\hat{\beta}) - f(\beta^0)\right)^\top \left(f(\hat{\beta}) - f(\beta^0)\right)\right] + \sigma^2 \\
&= \mathbb{E}\left[f(\hat{\beta}) - f(\beta^0)\right] \mathbb{E}\left[\left(f(\hat{\beta}) - f(\beta^0)\right)^\top\right] + \mathrm{Var}\left(f(\hat{\beta})\right) + \sigma^2 \\
&= \left|\left|\mathrm{Bias}\left(f(\hat{\beta})\right)\right|\right|^2 + \mathrm{Var}\left(f(\hat{\beta})\right) + \sigma^2.
\end{aligned}
$$

From the above discussion we observe that the total mean squared prediction error is the sum of the variance, the squared bias and a non-reducible error σ^2, which represents the randomness in the true model.

Especially when a model is designed with the purpose of forecasting, it needs to be able to generalize the relations beyond the scope of the dataset. Therefore the variables that are included in the model must be carefully chosen. Some frequently used measures to compare and assess models based on the *likelihood* and the number of variables included in the model are the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). We shall discuss measures after a quick introduction to likelihood functions.

Suppose the errors in our linear model are normally distributed. In that case, the likelihood for a model β is given by

$$L(\beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2}\left(\frac{y_i - X_i\beta}{\sigma}\right)^2\right) \tag{B.16}$$

The AIC is a measure for the quality of a model relative to another one, and is given by

$$\mathrm{AIC} = 2k - 2\ln(L(\beta, \sigma^2)), \tag{B.17}$$

where k is the number of parameters in the model and L the maximum value of the likelihood function of the model. The preferred model has the lowest AIC value. Thus it penalizes the inclusion of more parameters whilst rewarding an increase in the likelihood function. Hence the AIC finds a trade-off between goodness of fit and generalizability.

Similar to AIC is Bayesian Information Criterion (BIC). It is given by

$$\mathrm{BIC} = k \cdot \ln(n) - 2\ln \cdot (L), \tag{B.18}$$

where n is the number of observations in the dataset, and L the likelihood.

Another approach for selecting a model with the best predictive performance is to split a given data-set in a training and a test-sample. The model is then estimated on the training sample, and the predictive strength is tested on the prediction for the remaining test sample. Depending on the application and the model, there exist a myriad of adequate measures to determine the predictive strength of a model.

The (pseudo) R^2_{efron} method is a measure for the proportion of the variability in the response variables that is explained by the model. Let $\hat{y}_i$ denote the predicted responses. Then we have,

$$R^2_{\text{efron}} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i(\hat{\beta}))^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \tag{B.19}$$

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}. \tag{B.20}$$

Other measures to compare out-of-sample performance include Mean Absolute Error (MAE), Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE). They are calculated as follows.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^n |\hat{y}_i - y_i|, \tag{B.21}$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^n (\hat{y}_i - y_i)^2, \tag{B.22}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^n \left|\frac{\hat{y}_i - y_i}{y^i}\right|. \tag{B.23}$$

In this appendix we have seen a small selection of techniques that are used in regression and forecasting literature. Since many (introductory) statistics text books cover these topics extensively, we decided to include most of the statements in this appendix without proof.

Appendix C

KKT Conditions

In this appendix, we give the key aspects of a derivation of the Karush-Kuhn-Tucker (KKT) conditions for convex optimization problems, with an application to $\mathcal{L}^1$ penalized regression. Therefore we will proof the KKT-conditions for Lasso solutions that were given in Section 7.5. We shall also use central theorems from optimization theory that can be used to show the equivalence of the unconstrained and constrained Lasso problem (Lemma 7.4). Unless noted otherwise, the main results in this appendix are based on [11] and [13].

C.1 Convex Optimization

Definition C.1. (Convex Optimization) Let objective function f, and inequality constraint functions g_i, $i = 1, \dots, m$ be convex functions, equality constraints $h_j(x)$, $1, \dots, p$ affine functions, and $\mathcal{C} \subseteq \mathbb{R}^p$ a convex set. Then we say that,

$$\begin{array}{rlll} \text{minimize} & f(x) & & \\ \text{subject to} & g_i(x) \leq 0, & i = 1, \dots, p & \\ & h_i(x) = 0, & i = 1, \dots, m & \qquad x \in \mathcal{C}, \end{array}$$

is a convex optimization problem.

The convex optimization problem above is referred to as the *primal* problem in optimization theory. We assume that its domain $\mathcal{D} = \text{dom}(f) \cap_{i=1}^{p} \text{dom} g_i(x) \cap_{j=1}^{m}$ $\text{dom} h_j(x)$ is non-empty and denote its optimal solution by p^*. A closely related problem is the *Lagrange dual* problem.

Definition C.2. (Lagrange Dual Problem) The Lagrangian of the convex optimization problem is given by

$$L(x, \lambda, \nu) = f(x) + \sum_{j=1}^{p} \lambda_j g_j(x) + \sum_{i=1}^{m} \nu_i h_i(x). \tag{C.1}$$

The dual problem is given by

$$\begin{aligned} &\text{maximize} && \phi(\lambda, \nu) = \inf_{x \in \mathcal{C}} L(x, \lambda, \nu), \\ &\text{subject to} && \lambda \geq 0 \end{aligned}$$

We denote the optimal solution to the Lagrange dual problem by d^*. The difference with the primal solution, $p^* - d^*$ is referred to as the *duality gap*.

Definition C.3. (Duality) A convex optimization problem satisfies weak duality when $d^* \leq p^*$, and strong duality when $d^* = p^*$

Strong duality is a very useful property, and can for instance be used to show the equivalence of the constrained and unconstrained Lasso minimization problem. The condition in Definition **??** below is a sufficient condition for strong duality of a convex optimization problem. First we introduce the notation and concept of the *relative interior* of a set.

Let $\mathbf{aff}\,\mathcal{C}$ denote the *affine hull* of $\mathcal{C}$ (the set of all affine combinations of elements in $\mathcal{C}$). The relative interior $\mathbf{relint}\,\mathcal{C}$ is given by

$$\mathbf{relint}\,\mathcal{C} = \{x \in \mathcal{C} | \{y| \, ||y - x||_2 \leq r\} \cap \mathbf{aff}\,\mathcal{C} \subseteq \mathcal{C} \text{ for some } r > 0\}. \tag{C.2}$$

Definition C.4. (Slater's Condition) For a given convex minimization problem, there exists $x \in \mathbf{relint}\,\mathcal{C}$ such that

$$\begin{aligned} g_i(x) < 0, &\quad i = 1, \dots, p, \\ h_j(x) = 0, &\quad j = 1, \dots, m. \end{aligned}$$

Theorem C.5. *(Slater's Theorem) Consider a convex optimization problem of the form of Definition C.1. If Slater's condition (Definition C.4) holds, then strong duality holds.*

Proof. The proof of this theorem can be found in [11]. □

Strong duality allows us to proof that the KKT-conditions hold for any solution of the primal and dual problems. Below we show the intuition behind the derivation of these conditions.

When primal and dual solutions are attained and equal, we have that $d^* = p^*$. Let x^* be a primal optimal point and (λ^*, ν^*) the corresponding dual optimal point. Then we may write

$$
\begin{aligned}
f(x^*) &= \phi(\lambda^*, \nu^*) && \text{(C.3)}\\
&= \inf_{x \in \mathcal{C}} \left\{ f(x) + \sum_{j=1}^{p} \lambda_j^* g_j(x) + \sum_{i=1}^{m} \nu_i^* h_i(x) \right\} && \text{(C.4)}\\
&\leq f(x^*) + \sum_{j=1}^{p} \lambda_j^* g_j(x^*) + \sum_{i=1}^{m} \nu_i^* h_i(x^*) && \text{(C.5)}\\
&\leq f(x^*). && \text{(C.6)}
\end{aligned}
$$

The first equality C.3 is the definition of strong duality and the second equality C.4 holds by the definition of the Lagrangian dual. The inequality in C.5 follows from the fact that taking the infimum of $\phi(x)$ over x is always less or equal than the value of $\phi(x^*)$. Inequality C.6 follows from the fact that $h_i(x^*) = 0$, $\lambda_j^* \geq 0$, and $g_j(x^*) \leq 0$ for $i = 1, \ldots, m$ and $j = 1, \ldots, p$ by Definitions C.1 and C.2. We conclude that the inequality are in fact equalities by comparing the first and the last line.

From the above discussion, it follows that $\sum_{j=1}^{p} \lambda_j^* g_j(x^*) = 0$. Since $\lambda_j^* g_j(x^*) \leq 0$, $\forall j$ by the inequality constraint of the convex optimization problem, we know that $\lambda_j^* g_j(x^*) = 0$, for all $j = 1, \ldots, p$. This last expression is also commonly referred to as the *complementary slackness* condition.

Since x^* minimizes $L(x, \lambda^*, \nu^*)$ over x the derivative of L should be zero at x^*,

$$
\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i^* \nabla g_i(x^*) + \sum_{i=1}^{m} \nu_i^* \nabla h_i^*(x^*) = 0. \tag{C.7}
$$

Equation C.7 is referred to as the *stationarity* condition.

Combining the complementary slackness and stationarity conditions with feasibility conditions we obtain the KKT-conditions in Definition C.6 below.

Definition C.6. (KKT conditions) The Karush-Kuhn-Tucker conditions are given by

1. Stationarity: $\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i^* \nabla g_i(x^*) + \sum_{i=1}^{m} \nu_i^* \nabla h_i^*(x^*) = 0$
2. Complementary Slackness: $\lambda_j^* g_j(x^*) = 0$, $j = 1, \ldots, p$
3. Primal Feasibility: $g_i(x^*) \leq 0$, $h_i(x^*) = 0$, $i = 1, \ldots, m$
4. Dual Feasibility: $\lambda_j^* \geq 0$, $j = 1, \ldots, p$.

Theorem C.7. *(KKT Optimality) For a problem that satisfies strong duality, we have that x^* and λ^*, ν^* are primal and dual solutions if and only if x^* and λ^*, ν^* satisfy the KKT conditions.*

Proof. We first show that the KKT-conditions are sufficient, then we show that they are also necessary for primal and dual solutions to be equal.

- Sufficiency: Assume that x^* and λ^*, ν^* satisfy the KKT-conditions. Then

$$\begin{aligned}\phi(\lambda^*, \nu^*) &= L(x^*, \lambda^*, \nu^*) && \text{(by stationarity)}\\ &= f(x^*) + \sum_{j=1}^{p} \lambda_j^* g_j(x^*) + \sum_{i=1}^{m} \nu_i^* h_i(x^*) \\ &= f(x^*). && \text{(by complementary slackness, dual feasibility)}\end{aligned}$$

  It follows that x^* and $\lambda^*, \nu*$ are primal and dual optimal.

- Necessity: Assume that x^* and λ^*, ν^* are primal and dual optimal solutions. Then equations C.3-C.6 show that the KKT conditions are satisfied.

□

C.2 Lasso constrained optimization

The Lasso optimization problem is given by

$$\begin{aligned}\text{minimize} \quad & ||Y - X\beta||_2^2 \\ \text{subject to} \quad & \sum_{i=1}^{p} |\beta_j| - t \leq 0, \quad i = 1, \dots, p \\ & \beta \in \mathbb{R}^p.\end{aligned}$$

The objective function $||Y - X\beta||^2$ is strictly convex (by Lemma 7.15) and differentiable, the inequality constraint function $\sum_{j=1}^{p} |\beta_j|$ is convex and the equality constraint functions are zero and hence affine. Therefore, the lasso problem is a convex optimization problem as in Definition C.1.

The dual of the lasso problem is given by

$$d^* = \max_{\lambda} \phi(\lambda) = \max_{\lambda} \inf_{\beta} \frac{1}{2} ||Y - X\beta||_2^2 + \lambda \left(\sum_{j=1}^{p} |\beta_j| - t \right) \tag{C.8}$$

For $t > 0$ it is easy to see that the lasso problem satisfies Slater's condition. Put

$\beta = 0 \in \mathbf{relint}\, \mathbb{R}^p$, which gives the desired result $\sum_{j=1}^p |\beta_j| - t < 0$. Since the lasso optimization problem satisfies Slater's condition, it satisfies strong duality by Theorem C.5. This allows us to proof the equivalence of the Lasso problem in its primal and dual form.

Lemma C.8. *(Lasso Lagrange Dual) Assume that the minimization problem in Definition 7.3 is strictly feasible, then it is equivalent to*

$$d^* = \arg\min_{\beta} \frac{1}{2n} \sum_{i=1}^{n} (y_i - X_i\beta)^2 + \lambda^* ||\beta||_1. \tag{C.9}$$

Proof. For the proof we refer to Appendix F. □

By Theorem C.7 the lasso solution β^* satisfies the KKT-conditions:

1. Stationarity: $X^\top(Y - X\hat{\beta}) = \lambda s$, where

$$s(x) \in \begin{cases} \{1\} & x > 0, \\ [-1, 1] & x = 0, \\ \{-1\} & x < 0. \end{cases}$$

2. Complementary Slackness: $\lambda[\sum_{j=1}^p |\beta_j| - t] = 0$
3. Primal Feasibility: $\sum_{j=1}^p |\beta_j| \leq a$
4. Dual Feasibility: $\lambda \geq 0$.

C.3 Computation of Adaptive Lasso

The solution to the adaptively weighted version of Lasso minimization problem can be obtained by computing Lasso solutions for a scaled version of the design matrix by Proposition 8.2. The values for the tuning parameters λ and γ are obtained by finding the λ^* and γ^* with the minimal cross-validation error on the regularization path, as described in Section 8.1.3. The algorithm that can be used to obtain

Adaptive Lasso solution is described in Algorithm 2.

Data: design matrix X of size $n \times p$, response variable y of size $n \times 1$
Result: $\beta(\lambda_{\min}, \gamma_{\min})$ such that cross-validation error is minimal

initialize;
create folds $1, \ldots, k$;
create λ sequence $\lambda_1, \ldots, \lambda_s$;
create γ sequence $\gamma_1, \ldots, \gamma_t$;
create mean square error matrix of size $k \times s \times t$;

Find λ and γ such that the cross-validation error is minimal;
for $i \leftarrow 1$ **to** s **do**
  for $j \leftarrow 1$ **to** t **do**
    for $m \leftarrow 1$ **to** k **do**
      Construct training and test data from folds;
      $X_{\mathbf{train}} = X[-\text{fold}[m],]$, $X_{\mathbf{test}} = X[\text{fold}[m],]$;
      $y_{\mathbf{train}} = y[-\text{fold}[m]]$, $y_{\mathbf{test}} = y[\text{fold}[m]]$;
      compute $\beta^{\text{init}}_{\text{train}}(\lambda_i, \gamma_j)$ with normalized $X_{\mathbf{train}}$ and $y_{\mathbf{train}}$;
      weights $\leftarrow 1/\text{abs}\left(\beta^{\text{init}}_{\mathbf{train}}\left(\text{abs}\left(\beta^{\text{init}}_{\mathbf{train}}\right) > 0\right)\right)^{\gamma}$;
      mse$(i, j, m) \leftarrow$ mse for Adaptive Lasso solution for fold m, λ_i and γ_j;
    end
  end
end
mse = mean(mse , 3);
$\lambda_{\min} \leftarrow$ first coordinate of smallest value in mse;
$\gamma_{\min} \leftarrow$ second coordinate of smallest value in mse;

Compute Adaptive Lasso solution with λ_{min} and γ_{min};
compute β^{init} with normalized X and y;
weights $\leftarrow 1/\text{abs}(\beta^{\text{init}}(\text{abs}(\beta^{\text{init}}) > 0))^{\gamma_{\min}}$;
scale X by weights;
Compute $\beta(\text{weights}, \lambda_{\min}, \gamma_{\min})$ and $\beta_{\text{intercept}}$ with a lasso procedure;

Rescale the β estimate $\beta(\lambda_{\min}, \gamma_{\min}) \leftarrow \beta(\text{weights}, \lambda_{\min}, \gamma_{\min})/\text{weights}$;

Algorithm 2: AdaptiveLasso

Appendix D

Data Sources

In this appendix we provide the field names on the call reports that can be used to extract individual bank's time-series in Table D.1, the CAMELS criteria in Table D.2, and NCO rates for specific loan categories in Table D.3.

Table D.1: Large US Banks on Call Reports

Bank Name	Call Report ID number
JP Morgan Chase	852218
Bank of America	480228
Citigroup	476810
Wells Fargo	451965
Bank of New York Mellon	398668
HSBC North America Holdings	413208
PNC Financial Services Group	817824
Capital One Financial Corp	112837
TD Bank US Holding	497404
Morgan Stanley	1456501
Goldman Sachs	2182786

Table D.2: CAMELS Criteria on Call Reports

Financial Measure[1]	Call Report Fields
Total Assets	RCON2170
Equity	RIAD3210
Total Loans	RCON3360
Non Interest Expense	RIAD4093
Operating Income	RIAD4079
Net Income	RIAD4301
Liquid Assets[2]	RCFD0081 + RCFD0071 + RCONB987 + RCFDB989

[1] Total Loans for the specific loan categories can be derived from table D.3.
[2] Liquid assets include cash and balances due from depository institutions and federal funds.

Table D.3: Net Charge Off Fields on Call Reports

Loan Category	Valid for	Charge Off Fields	Recoveries Fields	Total Loans Fields
Loans to Depository Insitutions	1990Q2 - 2000Q4	RIAD4653 + RIAD4654	RIAD4663 + RIAD 4664	RCON1505 + RCON1517 + RCON1510
Loans to Depository Insitutions	2001Q1 - 2015Q2	RIAD4653 + RIAD4654	RIAD4663 + RIAD 4664	RCONB531 + RCONB534 + RCONB535
Real Estate Residential Loans	1990Q2 - 2001Q4	RIAD5413	RIAD5414	RCON5367 + RCON5368
Real Estate Residential Loans	2002Q2 - 2015Q2	RIADC234 + RIADC235	RIADC217 + RIADC218	RCON5367 + RCON5368
Real Estate HELOC Loans	1990Q2 - 2015Q2	RIAD5411	RIAD5412	RCON1797
Real Estate Construction & Land Development Loans	1990Q2 - 2007Q4	RIAD3582	RIAD3583	RCON1415
Real Estate Construction & Land Development Loans	2008Q1 - 2015Q2	RIADC891 + RIADC893	RIADC892 + RIADC894	RCONF158 + RCONF159
Real Estate Multi-Family Properties Loans	1990Q2 - 2015Q2	RIAD3588	RIAD3589	RCON1460
Real Estate Nonfarm Nonresidential Loans	1990Q2 - 2007Q4	RIAD3590	RIAD3591	RCON1480
Real Estate Nonfarm Nonresidential Loans	2008Q1 - 2015Q2	RIADC895 + RIADC897	RIADC896 + RIADC898	RCONF160 + RCONF161
Real Estate Farmland Loans	1990Q2 - 2015Q2	RIAD3584	RIAD3585	RCON1420
Commercial & Industrial Loans	1990Q2 - 2015Q2	RIAD4645	RIAD4617	RCON1763
Consumer Loans	1990Q2 - 2000Q4	RIAD4656 + RIAD4657	RIAD4666 + RIAD4667	RCON1975
Consumer Loans	2001Q1 - 2010Q4	RIADB514 + RIADB516	RIADB515 + RIADB517	RCONB538 + RCONB539 + RCON2011
Consumer Loans	2011Q1 - 2015Q2	RIADB514 + RIADK129 + RIADK205	RIADB515 + RIADK133 + RIADK206	RCONB538 + RCONB539 + RCONK137 + RCONK207
Agricultural Loans	1990Q2 - 2015Q2	RIAD4655	RIAD4665	RCON1590
Leases and Financing Receivables	1990Q2 - 2006Q4	RIAD4658 + RIAD4659	RIAD4668 + RIAD4669	RCON2182 + RCON2183
Leases and Financing Receivables	2007Q1 - 2015Q2	RIADF158 + RIADC880	RIADF187 + RIADF188	RCONF162 + RCONF163
Other Loans	2007Q1 - 2009Q4	RIAD4644	RIAD4628	RCON1563
Other Loans	2010Q1 - 2015Q2	RIAD4644	RIAD4628	RCONJ454 + RCONJ451 + RCON1545
Total Loans	1990Q2 - 2015Q2	RIAD4635	RIAD4605	RCON2122

Appendix E

Extra Results

In Tables E.1, E.2, E.3, and E.4 we display the stress identification results for 9-quarters ahead predictions on averaged NCO loan categories. For stress thresholds T_1, T_2, and T_3 we display the specificity, sensitivity, accuracy, and R^2_{count}, respectively. The results are displayed for the linear model selected and estimated by Adaptive Lasso, the RegARMA model, and the linear benchmark model.

Table E.1: Specificity Score for Stress Identification by Category

	Adaptive Lasso			RegARMA			Benchmark		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
HELOC									
RES			0.50			0.00			1.00
CLD	0.71	1.00	0.88	0.86	1.00	0.88	0.14	0.36	0.69
MF	0.82	0.74	0.80	0.82	0.79	0.76	0.09	0.42	0.72
NFNR	1.00	0.79	0.89	1.00	0.89	0.79	0.00	0.63	0.71
C&I	1.00	1.00	0.92	1.00	1.00	0.92	0.57	0.90	0.88
LEASE	0.65	1.00	0.92	0.65	1.00	0.92	0.35	0.71	0.92
CON	0.00	0.55	0.93	0.50	1.00	0.93	0.25	0.27	0.71
OTHER	0.46	1.00	0.95	0.46	1.00	1.00	0.38	0.50	0.75
AGRI	1.00	1.00	1.00	1.00	1.00	1.00	0.45	1.00	1.00
FARM	0.17	0.88	1.00	0.17	0.75	1.00	0.00	0.50	0.60
DEP	1.00	0.96	1.00	0.83	0.93	0.96	0.83	0.46	0.50

Table E.2: Sensitivity Score for Stress Identification by Category

	Adaptive Lasso			RegARMA			Benchmark		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
HELOC	0.68	0.36	0.21	0.64	0.50	0.36	0.68	0.68	0.61
RES	0.57	0.39	0.08	0.61	0.39	0.23	0.86	0.82	0.77
CLD	0.67	0.82	0.83	0.76	0.82	0.75	1.00	1.00	1.00
MF	0.82	0.89	0.67	0.82	1.00	1.00	1.00	1.00	1.00
NFNR	0.83	0.56		0.83	0.67		1.00	1.00	
C&I	0.71	0.88	1.00	0.71	1.00	1.00	0.64	0.88	0.50
CON	1.00	0.94	0.43	0.50	0.47	0.36	0.65	0.35	0.14
LEASE	0.91	1.00	0.75	0.73	0.86	0.75	0.73	0.43	0.25
OTHER	0.73	0.50	0.50	0.53	0.25	0.13	1.00	0.92	0.88
AGRI	0.12	0.00	0.00	0.06	0.11	0.00	0.06	0.00	0.00
FARM	0.45	0.15	0.00	0.41	0.15	0.17	0.95	0.75	0.61
DEP	0.06			0.19			0.88		

Table E.3: Accuracy Score for Stress Identification by Category

	Adaptive Lasso			RegARMA			Benchmark		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
HELOC	0.68	0.36	0.21	0.64	0.50	0.36	0.68	0.68	0.61
RES	0.57	0.39	0.11	0.61	0.39	0.21	0.86	0.82	0.79
CLD	0.68	0.89	0.86	0.79	0.89	0.82	0.79	0.75	0.82
MF	0.82	0.79	0.79	0.82	0.86	0.79	0.64	0.61	0.75
NFNR	0.89	0.71	0.89	0.89	0.82	0.79	0.64	0.75	0.71
C&I	0.86	0.96	0.93	0.86	1.00	0.93	0.61	0.89	0.82
CON	0.71	0.79	0.68	0.50	0.68	0.64	0.54	0.32	0.43
LEASE	0.75	1.00	0.89	0.68	0.96	0.89	0.50	0.64	0.82
OTHER	0.61	0.79	0.82	0.50	0.68	0.75	0.71	0.68	0.79
AGRI	0.46	0.68	0.79	0.43	0.71	0.79	0.21	0.68	0.79
FARM	0.39	0.36	0.36	0.36	0.32	0.46	0.75	0.68	0.61
DEP	0.46	0.96	1.00	0.46	0.93	0.96	0.86	0.46	0.50

Table E.4: R^2_{count} Score for Stress Identification by Category

	Adaptive Lasso			RegARMA			Benchmark		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
HELOC	0.00	−0.80	−2.67	0.00	0.00	−0.80	0.00	0.00	0.00
RES	0.00	−0.55	−5.25	0.00	−0.55	−1.75	0.00	0.00	0.25
CLD	0.36	0.79	0.71	0.50	0.79	0.64	0.14	0.36	0.69
MF	0.64	0.57	0.25	0.64	0.69	0.33	0.09	0.35	0.30
NFNR	0.77	0.38	0.00	0.77	0.55	0.00	0.00	0.56	0.00
C&I	0.71	0.88	0.67	0.71	1.00	0.67	0.42	0.70	0.29
CON	0.00	0.50	0.40	0.22	0.47	0.33	0.13	−0.12	0.00
LEASE	0.59	1.00	0.50	0.47	0.86	0.50	0.18	0.23	0.17
OTHER	0.35	0.50	0.44	0.30	0.25	0.13	0.38	0.47	0.54
AGRI	−0.15	0.00	0.00	−0.33	0.11	0.00	−0.83	0.00	0.00
FARM	−0.06	−0.64	−0.80	−0.20	−0.73	−0.15	0.00	0.31	0.35
DEP	−0.15	0.00	−	0.00	0.00	0.00	0.71	−0.15	0.00

Appendix F

Proofs of theorems

Proofs Concerning Convexity

Proof of Lemma 7.4. First we consider the primal problem,

$$p^* = \arg\min_{\beta} \frac{1}{2n}||Y - X\beta||_2^2 \quad \text{subject to } ||\beta||_1 \leq t. \tag{F.1}$$

The Lagrangian, the Lagrange dual, and the dual problem are given by,

$$\begin{aligned} L(\beta, \lambda) &= \frac{1}{2n}||Y - X\beta||_2^2 + \lambda(||\beta||_1 - t), \\ g(\lambda) &= \min_{\beta \in \mathbb{R}^p} L(\beta, \lambda), \\ d^* &= \max_{\lambda \geq 0} g(\lambda), \end{aligned}$$

respectively. To show equivalence of the constrained and unconstrained problem, we show both inclusions for their solution sets.

If the unconstrained problem is strictly feasible, we have that $t > 0$. For $t > 0$ it is easy to see that the primal problem satisfies Slater's condition. Put $\beta = 0 \in$ $\mathbf{relint}(\mathbb{R}^p)$, which gives the desired result $\sum_{j=1}^p |\beta_j| - y < 0$. Since the lasso optimization problem satisfies Slater's condition, it satisfies strong duality by theorem C.5. Then there exists $\lambda^* \geq 0$ such that any primal solution β^* minimizes

$$\frac{1}{2n}||Y - X\beta^*||_2^2 + \lambda^*(||\beta^*||_1 - t). \tag{F.2}$$

Hence, if β^* is a primal solution then (β^*, λ^*) is a minimizer of the unconstrained Lasso problem.

Conversely, we let $(\beta^*, \lambda*)$ be a solution of the unconstrained form. If $\lambda^* = 0$ then the unconstrained Lasso problem is equivalent to OLS regression, and the solution is also in the solution set of the constrained problem. Otherwise, by theorem C.7 the solution satisfies the KKT conditions, and hence the solution β^* satisfies $||\beta^*||_1 = t$ and is therefore in the solution set of the constrained Lasso problem.

We conclude that if the primal problem is strictly feasible then the two formulations have the same solutions. □

Proof of Proposition 7.5. From the complementary slackness condition of the KKT conditions, it follows that either $\lambda = 0$ or $\sum_{j=1}^{p} |\hat{\beta}_j| = t$. If $t < t_0$ if follows from primal feasibility and stationarity that $\lambda \neq 0$, and hence $\sum_{j=1}^{p} |\hat{\beta}_j| = t$. So $\hat{\beta}$ lies at the boundary of the constraint region. If $t \geq t_0$ then we have $\hat{\beta} = \arg\min_{\beta}\{||\beta||_1 : X^\top X\beta = X^\top Y\}$, with $\lambda = 0$. It is easily seen that $\hat{\beta}$ satisfies all KKT conditions and is thus an optimal solution. Because $\hat{\beta}$ is the solution to the normal equations it is an Ordinary Least Squares solution. □

Proof of Lemma 7.10. Let x be such that $f(x + ce_i) \geq f(x), \forall c, i$. We use the fact that a continuously differentiable function is convex if and only if $f(y) \geq f(x) + f'(x)(y - x)$, for all x, y. Hence we may write,

$$\begin{aligned}
f(y) - f(x) &\geq \nabla g(x)^\top (y - x) + \sum_{i=1}^{p} (h_i(y_i) - h_i(x_i)) \\
&= \sum_{i=1}^{p} \nabla_i g(x)(y_i - x_i) + (h_i(y_i) - h_i(x_i)) \geq 0.
\end{aligned}$$

Where the last inequality follows from the fact that each part of the sum is greater than or equal to zero by the assumption that $f(x + ce_i) > f(x)$ for all c, i. The reverse implication is trivial, since $f(x) = \min_{\beta \in \mathbb{R}^p} f(\beta)$ implies that $f(x) \leq f(y)$ for all y. □

Proof of Lemma 7.15.

i Suppose there exist two local minima x_1 and x_2 such that $f(x_1) \leq f(x_2)$. By strict convexity of f we have that

$$f(\alpha x_1 + (1 - \alpha)x_2) < \alpha f(x_1) + (1 - \alpha) f(x_2), \quad \alpha \in (0, 1). \tag{F.3}$$

Using strict convexity and the fact that for $\alpha > 0$ we have $\alpha f(x_1) \leq \alpha f(x_2)$, we may write

$$f(\alpha x_1 + (1-\alpha)x_2) < \alpha f(x_1) + (1-\alpha) f(x_2) \leq \alpha f(x_2) + (1-\alpha) f(x_2) = f(x_2). \tag{F.4}$$

Since α can be taken arbitrarily small, this contradicts the assumption that x_2 is a (local) minimum. Hence, there exists at most one minimization point.

ii Every convex function f on the Euclidean space is continuous. A closed and bounded set in Euclidian space is compact. The Extreme Value Theorem states that a continuous function attains its minimal and maximal values on a compact set.

iii Let $g = f + h$, we may write

$$g(\alpha x_1 + (1-\alpha)x_2) = f(\alpha x_1 + (1-\alpha)x_2) + h(\alpha x_1 + (1-\alpha)x_2) \tag{F.5}$$
$$< \alpha(f(x_1) + h(x_1)) + (1-\alpha)(f(x_2) + h(x_2)) \tag{F.6}$$
$$= \alpha g(x_1) + (1-\alpha)g(x_2), \tag{F.7}$$

where $\alpha \in (0,1)$. Where the strict inequality follows from the convexity of f and the strict convexity of h.

iv The result follows directly from the triangle inequality,

$$h(\alpha\beta_1 + (1-\alpha)\beta_2) = ||\alpha\beta_1 + (1-\alpha)\beta_2||_1 \leq \alpha||\beta_1||_1 + (1-\alpha)||\beta_2||_2. \tag{F.8}$$

v We rewrite,

$$\begin{aligned}
&||Y - \alpha X\beta_1 - (1-\alpha)X\beta_2||_2^2 \\
&\quad = \alpha^2||X\beta_1||_2^2 - 2\alpha^2(X\beta_1)(X\beta_2) + \alpha^2||X\beta_2||_2^2 - 2\alpha(X\beta_1)Y \\
&\qquad + 2\alpha(X\beta_1)(X\beta_2) + 2\alpha Y(X\beta_2) - 2\alpha||X\beta_2||_2^2 + ||y||_2^2 \\
&\qquad - 2Y(X\beta_2) + ||X\beta_2||_2^2.
\end{aligned}$$

Similarly we rewrite,

$$\begin{aligned}
&\alpha||Y - X\beta_1||_2^2 + (1-\alpha)||Y - X\beta_2||_2^2 \\
&\quad = \alpha||X\beta_1||_2^2 - 2\alpha(X\beta_1)Y + 2\alpha Y(X\beta_2) - \alpha||X\beta_2||_2^2 \\
&\qquad + ||Y||_2^2 - 2Y(X\beta_2) + ||X\beta_2||_2^2.
\end{aligned}$$

Combining the above results, we obtain

$$\begin{aligned}
&||Y - \alpha X\beta_1 - (1-\alpha)X\beta_2||_2^2 - \alpha||Y - X\beta_1||_2^2 - (1-\alpha)||Y - X\beta_2||_2^2 \\
&\quad = \alpha||X\beta_1||_2^2 - 2\alpha 2(X\beta_1)(X\beta_2) + \alpha^2||X\beta_2||_2^2 \\
&\qquad - \alpha||X\beta_1||_2^2 + 2\alpha(X\beta_1)(X\beta_2) - \alpha||X\beta_2||_2^2 \\
&\quad = \alpha(\alpha-1)||X\beta_1 - X\beta_2||_2^2.
\end{aligned}$$

Since $\alpha(\alpha-1)||X\beta_1 - X\beta_2||_2^2 < 0$ for all $X\beta_1 \neq X\beta_2$ and $\alpha \in (0,1)$, we find that $h(X\beta)$ is strictly convex. For $\beta_1 \neq \beta_2$ we have that $\alpha(\alpha-1)||X\beta_1 - X\beta_2||_2^2$ can also be equal to zero, if $X\beta_1 = X\beta_2$, hence $h(\beta)$ is convex.

vi If the columns of X are independent than there exist no $\beta_1 \neq \beta_2$ such that $X\beta_1 = X\beta_2$. Therefore equality between $||Y - \alpha X\beta_1 - (1-\alpha)X\beta_2||_2^2$ and $\alpha||Y - X\beta_1||_2^2 + (1-\alpha)||Y - X\beta_2||_2^2$ does not occur, and $h(\beta)$ is strictly convex.

□

Derivation of theorem 7.6

We closely follow [17] and [8]. Before we tackle the multi-dimensional case, we first look at Lemma F.1 below. In this lemma we consider a simpler, one-dimensional minimization problem.

Lemma F.1 (One-dimensional problem)**.** *Consider the minimization problem*

$$\arg\min_{\beta \in \mathbb{R}} l(\beta) = \arg\min_{\beta \in \mathbb{R}} \frac{1}{2}(z-\beta)^2 + J_\lambda(|\beta|), \tag{F.9}$$

where $J_\lambda(\cdot)$ is non-negative, non-decreasing, and differentiable on $(0,\infty)$. Assume that $-\beta - J'_\lambda(\theta)$ is strictly unimodal on $(0,\infty)$. Then the following holds,

i The solution exists, is unique, and is anti-symmetric ($\hat{\beta}(-z) = -\hat{\beta}(z)$).

ii The solution satisfies

$$\hat{\beta}(z) = \begin{cases} 0 & \text{if } |z| \leq p_0 \\ z - sgn(z)J'_\lambda(|\hat{\beta}(z)|) & \text{if } |z| > p_0 \end{cases} \tag{F.10}$$

where $p_0 = \min_{\beta \geq 0}\{\beta + J'_{\lambda(\beta)}\}$. Moreover $|\hat{\beta}(z)| \leq |z|$.

iii If $J'_\lambda(\cdot)$ is non-increasing then for $|z| \geq p_0$ we have $|z| - p_0 \leq |\hat{\beta}(z)| \leq |z| - J'_\lambda(|z|)$.

iv When $J'_\lambda(\beta)$ is continuous on $(0,\infty)$ then $\hat{\beta}(z)$ is continuous if and only if the minimum of $|\beta| + J'_\lambda(|\beta|)$ is attained at zero.

v If $J'_\lambda(|z|) \to 0$ as $z \to \infty$ then $\hat{\beta}(z) = z - J'_\lambda(|z|) + o(J'_\lambda(z))$.

Proof. We proof the statements of the lemma separately. The proof can also be found in [8].

1. The minimizer exists because $l(\beta) \to \infty$ as $|\beta| \to \infty$. Note that

$$l'(\beta) = \beta - z + J'_\lambda(\beta)\text{sgn}(\beta). \tag{F.11}$$

Without loss of generality we split the cases $z = 0$ and $z > 0$.

- $z = 0$: $\hat{\beta}(z) = 0$ is the unique minimizer
- $z > 0$: $\forall \beta > 0$ we have that $l(-\beta) > l(\theta)$ and hence $\hat{\beta}(z) > 0$. If $z < p_0$ then $l'(\beta) > 0$, so l is strictly increasing on $(0, \infty)$ and thus the solutions is unique. If $z \geq p_0$ then $\min l'(\beta) \leq 0$ and there are two possible zero crossings of $l(\beta)$. The larger one is the minimizer, because there the derivative is increasing. It follows that the solution is unique and satisfies $\hat{\beta}(z) = z - J'_\lambda(\hat{\beta}(z)) \leq z$.

2. If $|z| \leq p_0$ then $l'(\theta)$ is strictly positive on $(0, \infty)$ and hence $l(\beta)$ is strictly increasing. It follows that $\hat{\beta}(z) = 0$. When $|z| > p_0$ we saw for $z > 0$ that $\hat{\beta}(z) = z - J'_\lambda(\hat{\beta}(z)) \leq z$. Since for $z < 0$ we have that $\hat{\beta}(z) = z - J'_\lambda(\hat{\beta}(z)) \leq z$, the results follows directly
3. When $J'_\lambda(\cdot)$ is non-increasing it holds that $\hat{\beta}(z) \leq z - J'_\lambda(z)$. Let $\beta_0 = \min_{\beta \geq 0}\{\beta + J'_\lambda(\beta)\}$, then $\beta_0 < \hat{\beta}(z)$ for $z > p_0$. We have

$$J'_\lambda(\hat{\beta}(z)) \leq J'_\lambda(\beta_0) \leq \beta_0 + J'_\lambda(\beta_0) = p_0, \tag{F.12}$$

and the desired result follows.
4. Continuity for $z \neq p_0$ is guaranteed by the continuity of $\beta + J'_\lambda(\beta)$ on $(0, \infty)$. For $\hat{\beta}(z)$ is continuous at $z = p_0$ if and only if the minimum of $|\beta| + J'_\lambda(|\beta|)$ is attained at zero.
5. We know that $\hat{\beta}(z) = z - J'_\lambda(|\hat{\beta(z)}|)$. As $z \to \infty$, we must have $\hat{\beta} \to \infty$ and $J'_\lambda(\hat{\beta}(z)) \to 0$. The result follows.

□

From this lemma we can derive some results about conditions on the penalty function $J_\lambda(\beta)$ such that the properties of sparsity, continuity, and approximate unbiasedness are satisfied. From the fifth part of the lemma it is clear that a sufficient condition for the penalized regression estimator to be approximately unbiased is that $J_\lambda(|\beta|) \to 0$ as $\beta \to \infty$. In other words, if the penalty function is bounded by a constant, the resulting estimator is approximately unbiased. The second part of the lemma indicates that penalized regression has a *thresholding* property when $|z| \leq p_0$ and a *shrinkage* property when $|z| > p_0$. It follows that $\min_{\beta > 0}\{|\beta| + J'_\lambda(\beta)\} > 0$ is a sufficient condition for the sparsity property. The fourth part of the lemma shows that a necessary and sufficient condition for continuity is that the minimum of $|\beta| + J'_\lambda(\beta)$ is attained at 0.

The same reasoning can be employed to derive similar results for the penalized regression problem. Lemma 7.10 suggests that if the objective and the penalty function are convex then the (Adaptive) Lasso minimization problem can be solved coordinate-wise.

Lemma 7.10 implies that if a convex objective function with a convex penalty function is optimized along each coordinate axis iteratively, then the global minimum of the function is found, and vice versa. The convergence of minimizing the Lasso objective coordinate-wise were established in Section 8.1.3.

Proposition F.2. *Let $J_{\lambda_j}(\cdot)$ be a non-negative, non-decreasing, and convex function on $(0, \infty)$ and denote $\frac{1}{2n} X^\top X$ by $\hat{\mathcal{I}}$. Let $\tilde{\beta}$ be the minimizer of $Q(\beta)$ in equation 7.9, and recall that $\hat{\beta}$ is a minimizer of $||Y - X\beta||_2^2$. Then we have the following:*

1. *For the solution $\tilde{\beta}$ it holds that*

$$\tilde{\beta}_j = \begin{cases} 0 & \text{if } |\tilde{\beta}_j^*| \leq m_j, \\ \tilde{\beta}_j^* - \frac{1}{\hat{\mathcal{I}}_{jj}} J'_{\lambda_j}(|\tilde{\beta}_j|) sgn(\tilde{\beta}_j) & \text{if } |\tilde{\beta}_j^*| > m_j, \end{cases} \tag{F.13}$$

where

$$\tilde{\beta}_j^* = \hat{\beta}_j - \sum_{k=1}^{p} \frac{\hat{\mathcal{I}}_{jk}}{\hat{\mathcal{I}}_{jj}} (\beta_k - \hat{\beta}_k^{OLS}),$$
$$m_j = \min_{\theta > 0} \left\{ \theta + \frac{J'_{\lambda_j}(\theta)}{\hat{\mathcal{I}}_{jj}} \right\}.$$

2. *When $J'_{\lambda_j}(\beta_j)$ is continuous on $(0, \infty)$, the $\tilde{\beta}_j$ are continuous in $\hat{\beta}_j$ if and only if the minimum of $|\theta| + \frac{J'_{\lambda_j}(|\theta|)}{\hat{\mathcal{I}}_{jj}}$ is attained at zero for $j = 1, \ldots, p$.*

Proof. Below we give the proofs for both statements of the lemma.

1. Since $\widetilde{\beta}$ is a minimizer of $Q(\beta)$, it should hold that $\frac{\partial(Q(\beta))}{\partial \beta_j}|_{\widetilde{\beta}_j} = 0$by Lemma 7.10. We take the derivative of $Q(\beta)$ with respect to β_j. Since the solution is found by optimizing $Q(\beta)$ coordinate-wise, we fix β_k for $k \neq j$ and rewrite,

$$\begin{aligned} \frac{\partial Q(\beta)}{\partial \beta_j} &= \sum_{k=1}^{p} \hat{\mathcal{I}}_{jk}(\beta_k - \hat{\beta}_k) + J'_{\lambda_j}(|\beta_j|) sgn(\beta_j) \\ &= \hat{\mathcal{I}}_{jj} \left([\beta_j + J'_{\lambda_j}(|\beta_j|) sgn(\beta_j) / \hat{\mathcal{I}_{jj}}] - \beta_j^* \right) \end{aligned}$$

Note that $\hat{\mathcal{I}}_{jj} = (X^\top X)_{jj} = \sum_{i=1}^{p} X_{i,j}^2 \geq 0$. To find the minimum we equate the derivative to zero. We distinguish two cases:

- $|\beta_j^*| < m_j$: $\widetilde{\beta}_j = 0$ because otherwise, if $|\widetilde{\beta_j}| > 0$ we have that $\frac{\partial(Q(\widetilde{\beta}))}{\partial \widetilde{\beta}_j} \neq 0$, a contradiction.

- $|\beta_j^*| \geq m_j$: We set the derivative to zero and solve the equation for β_j, the result in equation F.13 follows by rewriting.

2. By continuity of J'_{λ_j} and the result of equation F.13, we have that $\widetilde{\beta}_j$ is continuous in $\hat{\beta}_j$ when $|\widetilde{\beta}_j^*| \neq m_j$. It remains to show continuity for the case that $|\tilde{\beta}_j^*| = m_j$. For $\widetilde{\beta}$ and $|\beta_j^*| = m_j$ we have that

$$\frac{\partial Q(\tilde{\beta})}{\partial \beta_j} = 0 \implies [\tilde{\beta}_j + J'_{\lambda_j}(|\tilde{\beta}_j|)sgn(\tilde{\beta}_j)/\hat{\mathcal{I}_{jj}}] = \beta_j^* = \min_{\theta>0}\left\{\theta + \frac{J'_{\lambda_j}(\theta)}{\hat{\mathcal{I}}_{jj}}\right\},$$

It follows that $\tilde{\beta}_j = 0$ if and only if the minimum of $|\theta| + \frac{J'_{\lambda_j}(|\theta|)}{\hat{\mathcal{I}}_{jj}}$ is attained at zero. In that case the solutions are continuous.

□

The results that we derived can be used to give some guidance on how to choose a penalty function.

(Sketch of proof of Theorem 7.6). We give an outline of part of the proof that appears in [17]. By assumption the conditions of proposition F.2 hold. The first part of the proposition implies that if coefficients are set to zero, it should hold that $m_j > 0$. In that case we have that $J'_{\lambda_j}(|\cdot|)$ is positive around zero. Since $J_{\lambda_j}(\cdot)$ is a non-decreasing function on $(0, \infty)$, the function $J_{\lambda_j}(|\cdot|)$ on $\mathbb{R}$ attains its minimum at zero. Hence $J_{\lambda_j}(|\cdot|)$ attains its minimum at zero, and has positive derivative is positive around zero, hence it is non-differentiable at the origin. The second part of Proposition F.2 gives necessary and sufficient conditions for continuity of the resulting estimator. □

Other Proofs

Proof of Lemma 7.9. We compute the Taylor expansion of the $\mathcal{L}^2$ loss. The derivatives of $||Y - X\beta||_2^2$ are given by:

$$\frac{\partial}{\partial \beta}\left[||Y - X\beta||_2^2\right] = -2X^\top Y + 2X^\top X\beta,$$
$$\frac{\partial^2}{\partial \beta^2}\left[||Y - X\beta||_2^2\right] = 2X^\top X,$$
$$\frac{\partial^k}{\partial \beta^k}\left[||Y - X\beta||_2^2\right] = 0, \quad k \geq 3.$$

Since $\hat{\beta}$ is a minimizer of $||Y-X\beta||_2^2$ by assumption, it follows that $\frac{\partial}{\partial\beta}\left(||Y-X\hat{\beta}||_2^2\right) = 0$. The Taylor expansion of $||Y-X\beta||_2^2$ in $\hat{\beta}$ is given by:

$$\begin{aligned}
||Y-X\beta||^2 &= ||Y-X\hat{\beta}||^2 + \left(-2X^\top Y + 2X^\top X\hat{\beta}\right)(\beta-\hat{\beta}) + (\beta-\hat{\beta})^\top X^\top X(\beta-\hat{\beta}) \\
&= ||Y-X\hat{\beta}||^2 + (\beta-\hat{\beta})^\top X^\top X(\beta-\hat{\beta}).
\end{aligned}$$

Since the first term is constant with respect to β, the result follows directly by plugging in the above in equation 7.8. $\square$

Sketch of proof of Theorem 7.13. The sketch given here follows the proof in [16]. We start with the proof for asymptotic normality (property 2), and finish with the proof for sign consistency (property 1).

- Asymptotic Normality: Let β^0 be the true model and write $\beta = \beta^0 + u/\sqrt{n}$. Consider $\hat{u}^{(n)} = \arg\min_u \Psi_n(u)$ where,

$$\begin{aligned}
\Psi_n(u) &= \arg\min_u \left|\left|Y - \sum_{j=1}^p x_j\left(\beta_j^0 + \frac{u_j}{\sqrt{n}}\right)\right|\right|^2 + \lambda_n \sum_{j=1}^p \hat{w}_j \left|\beta_j^0 + \frac{u_j}{\sqrt{n}}\right| \\
\Psi_n(u) - \Psi_n(0) &= u^\top \frac{1}{n}(X^\top X)u - 2\frac{\epsilon^\top X}{\sqrt{n}} + \frac{\lambda_n}{\sqrt{n}}\sum_{j=1}^p \hat{w}_j\sqrt{n}\left(\left|\beta_j^0 + \frac{u_j}{\sqrt{n}}\right| - |\beta_j^0|\right)
\end{aligned}$$

  Observe that $\hat{\beta}^{0(n)} = \beta^0 + \frac{\hat{u}^{(n)}}{\sqrt{n}}$ so $\hat{u}^{(n)} = \sqrt{n}(\hat{\beta}^{0(n)} - \beta^0)$.
  By assumption $\frac{1}{n}X^TX \to C$ and $\frac{\epsilon^\top X}{\sqrt{n}} \xrightarrow{d} W = \mathcal{N}(0,\sigma^2 C)$ (since the errors $\epsilon_i = Y_i - X_i\beta \sim \mathcal{N}(0,\sigma^2)$), where C is a positive definite matrix). Hence, we consider the limiting behavior of the last term. If $\beta^0 \neq 0$, then

$$\hat{w}_j \xrightarrow{p} |\beta_j^0|^{-\gamma}, \tag{F.14}$$

$$\sqrt{n}\left(\left|\beta_j^0 + \frac{u_j}{\sqrt{n}}\right| - |\beta_j^0|\right) \xrightarrow{p} 0. \tag{F.15}$$

  By Slutsky's theorem the third term converges to 0. For $\beta_j^0 = 0$ a similar result can be derived. Then, using Slutsky's theorem again, it follows that $\forall u$,

$$\Psi_n(u) - \Psi_n(0) \xrightarrow{d} \begin{cases} u_S^\top C_{11}u_S - 2u_S^\top W_S & \text{if } u_j = 0\,\forall j \notin S \\ \infty & \text{otherwise} \end{cases} \tag{F.16}$$

  Ìt can be shown that $\Psi_n(u) - \Psi_n(0)$ is convex and has the unique minimum $(C_{11}W_S, 0, 0, ..)^\top$, then using convergence results in [7] it can be derived that for the limiting behavior of $\hat{u}^{(n)}$ it holds that,

$$\hat{u}_S^{(n)} \xrightarrow{d} C_{11} W_S \qquad \hat{u}_{S^C}^{(n)} \xrightarrow{d} 0. \tag{F.17}$$

Since $W_S = \mathcal{N}(0, \sigma^2 C_{11})$, it follows that $\hat{u}_S^{(n)} \xrightarrow{d} \mathcal{N}(0, \sigma^2 C_{11}^{-1})$.

- Sign Consistency: By the asymptotic normality we now that $\forall j \in S$

$$\hat{\beta}_j^{(n)} \xrightarrow{p} \beta_j^0. \tag{F.18}$$

This implies that $\mathbb{P}(j \in S_n) \to 1$. Hence, it suffices to show that $\forall k \notin S$ we have $\mathbb{P}(k \in S_n) \to 0$. From the KKT conditions it follows that $2X_k^\top (y - X\hat{\beta}^{(n)}) = \lambda_n \hat{w}_k$. It can be shown that

$$\mathbb{P}\left(2X_k^\top (y - X\hat{\beta}^{(n)}) = \lambda_n \hat{w}_k\right) \to 0. \tag{F.19}$$

Since $\mathbb{P}(k \in S_n) \leq \mathbb{P}(2X_k^\top (y - X\hat{\beta}^{(n)})$, the results follows directly.

$\square$

References

[1] Roger Koenker and Jr. Gilbert Basset. "Regression Quantiles". In: *Econometrica* 46.1 (1978), pp. 33–50.

[2] Robert Tibshirani. "Regression Shrinkage and Selection via the Lasso". In: *Journal of the Royal Statistical Society* 58.1 (1996), pp. 267–288.

[3] Chi-Chun Liu, Stephen G. Ryan, and James M. Wahlen. "Differential Valuation Implications of Loan Loss Provisions across Banks and Fiscal Quarters". In: *The Accounting Review* 72.1 (1997), pp. 133–146.

[4] Beverly Hirtle and Jose A. Lopez. "Supervisory Information and the Frequency of Bank Examinations". In: *Federal Reserve Bank of the US, New York, Economic Policy Review* 5 (1999), pp. 1–20.

[5] Richard Weber. "Time Series". University of Cambridge, Lecture Notes Statistical Laboratory. 1999.

[6] Anil K. Kashyap and Jeremy C. Stein and. "What do a million observations of banks say about the transmission of monetary policy?" In: *American Economic Review* 90.3 (2000), pp. 407–428.

[7] Keith Knight and Wenjiang Fu. "Asymptotics for Lasso-Type Estimators". In: *The Annals of Statistics* 28.5 (2000), pp. 1356–1378.

[8] Anestis Antoniadis and Jinqing Fan. "Regularization of Wavelet Approximations". In: *Journal of the American Statistical Association* 96.455 (2001), pp. 939–955.

[9] Jianqing Fan and Runze Li. "Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties". In: *Journal of the American Statistical Association* 96 (2001), pp. 1348–1360.

[10] Victor Zarnowitz and Ataman Ozyildirim. "Time Series Decompositions and Measurement of Business Cycles, Trends and Growth Cycles". 2002.

[11] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004. ISBN: 0521833787.

[12] René Doff. *Economic Capital en risicobeheer bij banken*. Nibesvv, 2004. ISBN: 9055162086.

[13] Etienne de Klerk, Kees Roos, and Tamas Terlaky. “Nonlinear Optimization”. University of Waterloo, Lecture Notes. 2005.

[14] Jian Huang, Shuangge Ma, and Cun-hui Zhang. *Adaptive Lasso for sparse highdimensional regression.* Tech. rep. University of Iowa, 2006.

[15] Peng Zhao and Bin Yu. “On Model Selection Consistency of Lasso”. In: *Journal of Machine Learning Research* 7 (2006), pp. 2541–2563.

[16] Hui Zou. “The Adaptive Lasso and Its Oracle Properties”. In: *Journal of American Statistical Association* 101.476 (2006), pp. 1418–1429.

[17] Yiyun Zhang and Runze Li. “Iterative Conditional Maximization Algorithm for Nonconcave Penalized Likelihood”. In: *Nonparametric Statistics and Mixture Models: A Festschrift in Honor of Thomas P Hettmansperger* (2008).

[18] Sara van de Geer and Peter Buhlmann. “On the conditions used to prove oracle results for the lasso”. In: *Electronic Journal of Statistics* 3 (2009), pp. 1360–1392.

[19] Trevor Hastie, Rob Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data mining, Inference and Prediction.* Springer Series in Statistics, 2009.

[20] Yingying Li and Stanley Osher. “Coordinate descent optimization for l1 minimization with application to compressed sensing; a greedy algorithm”. In: *Inverse Problems and Imaging* 3.3 (2009), pp. 487–503.

[21] Mario Quagliariello. “Macroeconomic stress-testing: definitions and main components”. In: *Stress-testing the Banking System.* Cambridge University Press, 2009.

[22] Jerome Coffinet and Surong Lin. “Stress Testing Banks’ Profitability: The Case of French Banks”. In: *Documents de Travail, Banque de France* 306 (2010).

[23] David Pardoe and Peter Stone. “Boosting for Regression Transfer”. In: *Proceedings of the 27th International Conference on Machine Learning (ICML).* 2010.

[24] Xiaoiao Shi et al. “Predictive Modeling with Heterogenous Sources”. in: Proceedings of the 2010 SIAM International Conference on Data Mining. 2010.

[25] Peter Bühlmann and Sara van de Geer. *Statistics for High-Dimensional Data.* Springer Series in Statistics, 2011.

[26] Peter Bühlmann and Sara van de Geer. *Statistics for High-Dimensional Data: Theory, Methods and Applications.* Springer, 2011. ISBN: 0172-7397. DOI: `10.1007/978-3-642-20192-9`.

[27] Jianqing Fan, Jinchi Lv, and Lei Qi. “Sparse High-Dimensional Models in Economics”. In: *Annual Review of Economics* 3.1 (2011), pp. 291–317.

[28] Yongdai Kim. “On weak and strong oracle properties”. Notes of workshop Current Challenges in Statistical Learning. 2011.

[29] Vahid Behbood. “Fuzzy Transfer Learning for Financial Early Warning System”. Phd thesis submitted at University of Technology, Sydney. 2012.

[30] Luca Guerrieri and Michelle Welch. “Can Macro Variables Used in Stress Testing Forecast the Performance of Banks?” In: *FEDS Working Paper* 7.2012-49 (2012). DOI: `http://dx.doi.org/10.2139/ssrn.2192650`.

[31] Anwarul Kabir and Suman Dey. “Peformance Analysis through CAMEL rating: A Comparitive Study of Selected Private Commercial Banks in Bangladesh”. In: *Journal of Politics and Governance* 1.2 (2012), pp. 16–25.

[32] Rajen Shah. “The Lasso: Variable Selection, Prediction and Estimation”. Notes on talk at Statistics Reading Group 2012. 2012.

[33] Rongning Wu and Qin Wang. “Shrinkage Estimation for linear regression with ARMA errors”. In: *Journal of Statistical Planning and Inference* 142 (2012), pp. 2136–2148.

[34] Lee Dicker, Baosheng Huang, and Xihong Lin. “Variable Selection and Estimation with the Seamless-L0 Penalty”. In: *Statistica Sinica* 23 (2013), pp. 929–962.

[35] Michael Duane, Til Schuermann, and Peter Reynolds. “Stress Testing Bank Profitability”. In: *Journal of Risk Management in Financial Institutions* 7.1 (2013), pp. 72–84.

[36] Camila Epprecht, Dominique Guegan, and Álvaro Veiga. “Comparing variable selection techniques for linear regression: LASSO and Autometrics”. In: *Documents de Travail du Centre d’Economie de la Sorbonne* (2013).

[37] Wolfgang Hess et al. *Using Lasso-Type Penalties to Model Time-Varying Covariate Effects in Panel Data Regressions - A Novel Approach Illustrated by the ’Death of Distance’*. IFN Working Paper No. 961. 2013.

[38] Malene Kallestrub-Lamb, Anders Bredahl Kock, and Johannes Tang Kristensen. “Lassoing the Determinants of Retirement”. In: *CREATES Research Papers* 2013-21 (2013).

[39] Ryan J. Tibshirani. “The lasso problem and uniqueness”. In: *Electronic Journal of Statistics* 7 (2013), pp. 1456–1490. DOI: `10.1214/13-ejs815`.

[40] Aad W. van der Vaart. “Time Series”. Universiteit Leiden, Lecture Notes version 242013. 2013.

[41] Xinpeng Wang. “Use SAS/ETS to Forecast Credit Loss for CCAR”. MWSUG 2013 Conference Proceedings, September 22-24. 2013.

[42] Basel Committee on Banking Supervision. “A Sound Capital Planning Process: Fundamental Elements”. In: (2014).

[43] Francisco Covas, Ben Rump, and Egon Zakrajšek. “Stress-Testing U.S. Bank Holding Companies: A Dynamic Panel Quantile Regression Approach”. In: *International Journal of Forecasting* 663 (2014).

[44] Jerome Friedman et al. *glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models*. R-package version 2.0-2. 2014.

[45] Jochen Garcke and Thomas Vanck. “Importance Weighted Inductive Transfer Learning for Regression”. In: *Machine Learning and Knowledge Discovery in Databases*. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2014, pp. 466–481. DOI: `10.1007/978-3-662-44848-9_30`.

[46] Tim Harford. “Big Data: Are we making a big mistake?” In: *Significance* 11.5 (2014), pp. 14–19. DOI: `10.1111/j.1740-9713.2014.00778.x`.

[47] Beverly Hirtle et al. “The Capital and Loss Assesment under Stress Scenarios (CLASS) Model”. In: *Federal Reserve Bank of New York Staff Report* 30.3 (2014), pp. 691–713.

[48] Anders Bredahl Kock and Timo Teräsvirta. “Forecasting performances of three automated modelling techniques during the economic crisis 2007-2009”. In: *International Journal of Forecasting* 30 (2014), pp. 616–631.

[49] Clifford Lam and Pedro Souza. *Regularization for Spatial Panel Time Series Using the Adaptive LASSO*. STICERD - Econometrics Paper Series. Suntory, Toyota International Centres for Economics, and Relate Disciplines, LSE, 2014.

[50] Jimuta Naik. “ Beginning of the Early Banking Industry in Mesopotamia Civilization from 8th Century B.C.E. ” In: (2014). DOI: `http://dx.doi.org/10.2139/ssrn.2377309`.

[51] Jurgen A. Doornik and David F. Hendry. “Statistical model selection with big data”. In: *Cogent economics & finance* (2015). DOI: `http://dx.doi.org/10.1080/23322039.2015.1045216`.

[52] Rabobank Group. *Interim Report 2015*. 2015.

[53] Trevor S. Harris, Urooj Khan, and Doron Nissim. “The Expected Rate of Credit Losses on Banks’ Loan Portfolios”. Columbia Business School Research Paper No. 15–21. 2015.

[54] Pavel Kapinos and Oscar A. Mitnik. “A Top-Down Approach to the Stress-testing of Banks”. In: (2015). DOI: `http://dx.doi.org/10.2139/ssrn.2469425`.

[55] Todd Knoop. *Business Cycle Economics*. Praeger, 2015. ISBN: 1440831742.

[56] Wouter J. den Haan, Steven W. Sumner, and Guy M. Yamashiro. "Construction of Aggregate and Regional Bank Data Using the Call Reports: Data Manual". Unpublished Manuscript, University of Amsterdam (2002).

[57] Anna Mikusheva. "course materials for 14.384 Time Series Analysis". Downloaded on 08-10-2015. Fall 2007.