Charles University

Faculty of Social Sciences
Institute of Economic Studies

MASTER'S THESIS

Satellite Model Accuracy in Bank Stress Testing

Author: **Bc. Filip Hamáček**

Supervisor: **Mgr. Petr Polák MSc.**

Academic Year: **2018/2019**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain a different or the same degree.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, January 4, 2019

Signature

Acknowledgments

I would like to express gratitude to my thesis supervisor, Mgr. Petr Polák MSc., for providing me the opportunity to write about the topic that interested me. I am also grateful to everyone who supported me while I was writing, to my brother that gave me advices from different perspective and to everyone that gave me advice about the correct required PDF format.

Abstract

This thesis is dealing with credit risk satellite models in Czech Republic. Satellite model is a tool to predict financial variable from macroeconomic variables and is useful for stress testing the resilience of the banking sector. The aim of this thesis is to test accuracy of prediction models for Probability of Default in three different segments of loans - Corporate, Housing and Consumer. Model currently used in Czech National Bank is fairly unchanged since 2012 and its predictions can be improved. This thesis tests accuracy of the original model from CNB by developing new models using modern techniques, mainly by model combination methods: Bayesian Model Averaging (currently used in European Central Bank) and Frequentist Model Averaging. Last approach used are Neural Networks.

JEL Classification	C53, E17, E51, E58
Keywords	satellite model, credit risk, Czech economy, model combination

Author's e-mail	`hamacekf@gmail.com`
Supervisor's e-mail	`polakpet@gmail.com`

Abstrakt

Tato práce se zabývá satelitními modely kreditního rizika v České republice. Satelitní model je nástroj pro odhadování finančních proměnných z makroekonomických proměnných. Takový model je užitečný pro stres testování odolnosti bankovního sektoru. Cílem této práce je testování přesnosti modelu pro pravděpodobnost selhání ve třech segmentech úvěrů - korporátní úvěry, úvěry na bydlení a spotřebitelské úvěry. Současně využívaný model v České Národní Bance je z roku 2012 a jeho výkonnost není ideální. Tato práce testuje přesnost tohoto modelu vytvořením alternativních modelů pomocí moderních metod, především metod kombinujících modely: Bayesovské kombinování modelů (využívané v Evropské Centrální Bance) a Frekvenční kombinování modelů. Poslední metodou jsou Neuronové sítě.

Klasifikace JEL	C53, E17, E51, E58
Klíčová slova	satelitní model, kreditní riziko, Česká ekonomika, kombinování modelů

E-mail autora	hamacekf@gmail.com
E-mail vedoucího práce	polakpet@gmail.com

Contents

List of Tables

List of Figures

Acronyms

CNB Czech National Bank

ECB European National Bank

FMA Frequentist Model Averaging

BMA Bayesian Model Averaging

NN Neural Network

RMSE Root Mean Square Error

MAE Mean Absolute Error

Master's Thesis Proposal

Author	Bc. Filip Hamáček
Supervisor	Mgr. Petr Polák MSc.
Proposed topic	Satellite Model Accuracy in Bank Stress Testing

Motivation Starting in 1990s and later in 2000s the new phenomen of testing adverse scenarios in banking system has been developed. The satellite stress models in general put into relation macroeconomic variables with financial variables to test how external changes can affect stability of individual institution or the whole sector. The proposed topic is aiming to develop new model for testing resilience of Czech banking sector.

From simple stress models at the beginning (Blaschke, 2001), more complicated models have been developed (Cihak, 2005). The original IMF and CNB methodology is provided in Cihak (2007). However the main deficiencies were discovered during the financial crisis of 2007 – 2008 when most of the tests did not reveal high risk. According to the models the system should have remained stable. The new approach was developed after the crisis, however since 2012 there was not much of a progress in stress testing (especially in Czech republic). The main property of after crisis tests is conservative calibration, which assumes rather pessimistic scenarios. The goal is not to underestimate the risk, so the system has the conservative buffer. The current stress test methodology framework of CNB is described in Gersl (2012). Current stress models in CNB work in a way that the endogenous shock is introduced into NiGEM model (model for global economy to produce trajectories of foreign variables) and to g3 prediction model (DSGE model used in CNB for domestic variables). Predictions from these models are then introduced into satellite models that produce results, such as probability of default or property prices. The framework of ECB can be found in Dees (2017).

The aim of this thesis is to develop new credit risk satellite models to test for bank-stress testing in the Czech republic, test their predictions and compare performance to current CNB methodology, which is unchanged since 2012. Models from CNB are being re-estimated each period to obtain new coefficients for new predictions, however the model framework is old. At first, other (more complicated and modern) frameworks can be used to current CNB models that are used to model financial variables, for example models used in ECB. These models can perform better than ARIMAX and ARDL used in CNB. Second of all, the new models should try to more investigate variables that are connected with unemployment as those variables do not play that crucial role in Czech frameworks, but they are very important in US frameworks. Third of all, it might be convenient to try the combination of more models (Papadopoulos, 2016). At first stage we can choose surviving candidate models, then these models are combined through weights to obtain the final model, which covers wide range of stress scenarios.

Hypotheses

Hypothesis #1: 1. Other models can perform better than ARIMAX used by Czech National Bank.

Hypothesis #2: 2. Combining more than one model can yield better results than just one model.

Hypothesis #3: 3. Unemployment has crucial role in Satellite models and should be used in CNB models.

Methodology Current CNB framework is based on ARIMA models. Following recent development of ECB macro stress tesing, this thesis will compare performace of new methodods and test if the prediction power can be significantly improved. Following methodologies will be compared: ARIMA, BMA, FMA, Neural networks. Outomes of the models will be compared with real data from the economy and with predictions published by CNB. The main difference with respect to CNB framework for the satellite models will be the method of Artificial Neural Networks (ANN). This method is based on the fact that original input variables affect in nonlinear way other variables (hidden neurons) stacked in hidden layer (or more hidden layers) where these hidden neurons affect in nonlinear way our dependent variable. As with increasing number of hidden neurons and hidden layers the number of parameters to be estimated is exponentially increasing, hence it is convenient to have rather smaller networks. Parameters estimation in most of ANN techniques is done by nonlinear least squares (NLS) or conditional maximum likelihood (ML). The vector of parameters will be estimated by Broyden-Fletcher-Goldfarb-Shannon (BFGS) algorithm or by Levenberg-Marquardt algoritm as suggested in Rech (2002). For the modelling, several procedures will be used. First one is Early Stopping, which is based on dividing dataset into 3 subsets, training set for estimating parameters, validation set and test set. Another technique is Prunning, choosing the smallest, good forecasting model based on different methods, such as Information Criterion Prunning, Cross-Validation Prunning or Interactive Prunning. Simmilar technique is Regularization where the goal is to find balance between number of parameters and goodness of fit by penalizing large models. The last technique that will be used is Statistical approach (SA). As the adverse scenarios are used in stress models, simulated events have not been realized. To compare quasi out-of-sample performance of the original Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) or Autoregressive Distributed Lag (ARDL) model with models produced by ANN, these metrics can be used: MAPE, WMAPE, MAE and most of the time RMSE. For the model combining the Bayesian Model Averaging (BMA) and the Frequentist Model Averaging (FMA) will be used. Both models are allowing to combine more models and capture wider spectrum of scenarios. These frameworks are assigning weights to candidate models, for older BMA choosing prior probabilities has to be set. The newer FMA on the other hand does not require any priors, weights and estimators are determined by data.

Expected Contribution Current methodology of CNB for satellite models is old, so the new development is needed. The thesis can obtain the most recent models that fit better the new data. CNB methodology can be improved by introducing new modern estimation methods. The whole framework can be changed in reflection of the newest irregular events (zero interest rates, deflation, ...). Combining models can be more suitable than just searching for one true model. Models with more influence of variables related to unemployment can be more accurate and they can cover wider space of possible scenarios. This thesis will be unique in a way that this approach has never been done yet for Czech system in any paper. Main contribution should be finding best-practice for stress-testing using current availible methods.

Outline

1. Introduction, motivation
2. Literature Overview
3. Methodology, theoretical overview
4. Empirical analysis of different estimation frameworks
5. Introducing variables connected to unemployment into the models
6. Presenting and comparing results
7. Conclusion, Discussion

Core bibliography

DEES, Stephane, Henry JEROME a Martin REINER. Stress-Test Analytics for Macroprudential Purposes in the euro area. European Central Bank, 2017.

GERSL, Adam, Petr JAKUBIK, Tomas KONECNY a Jakub SEIDLER. Dynamic Stress Testing: The Framework for Testing Banking Sector Resilience Used by the Czech National Bank. Czech National Bank, 2012.

HENRY, Jerome. A macro stress testing framework for assessing systemic risks in the banking sector. European Central Bank, 2013.

PAPADOPOULOS, George, Savas PAPADOPULOS a Thomas SAGER. Credit risk stress testing for EU15 banks: a model combination approach. Bank of Greece, 2016.

CIHAK, Martin a Jaroslav HERMANEK. Stress Testing the Czech Banking System: Where Are We? Where Are We Going?. Czech National Bank, 2005.

CIHAK, Martin. Introduction to Applied Stress Testing. International Monetary Fund, 2007.

FOGLIA, Antonella. Stress testing credit risk: a survey of authorities' approaches. Bank of Italy, 2008.

BLASCHKE, Winfrid, Matthew T. JONES, Giovanni MAJNONI a Soledad Martinez PERIA. Stress Testing of Financial Systems: An Overview of Issues, Methodologies, and FSAP Experiences. International Monetary Fund, 2001.

RECH, Gianluigi. Forecasting with artificial neural network models. Stockholm, Sweden: Department of Economic Statistics, Stockholm School of Economics, 2002.

Author Supervisor

Chapter 1

Introduction

Starting in 1990s and later in 2000s the new phenomen of testing adverse scenarios in banking system has been developed. The satellite stress models in general put into relation macroeconomic variables with financial variables to test how external changes can affect stability of individual institution or the whole sector. The topic is aiming to develop new model for testing resilience of Czech banking sector, in particular forecasting model of Probability of Default of subjects in given segment. This model is useful in foreseeing crisis periods, testing resilience of the banking sector when introducing shocks or constructions of adverse scenarios. The financial crisis of 2008 has uncovered problems in stress testing until that period, because most of the models expected the banking sector to remain stable by underestimating the credit risk. Current methodology of Czech National Bank is developed after the crisis period, but it is fairly unchanged since 2012 and uses old techniques, likely not predicting well possible incoming recession periods.

The objective of this thesis is to develop alternative models using modern techniques and compare them to the model of the Czech National Bank to test accuracy of the alternative models as well as the original model, which is re-estimated on new available data. The whole framework can be changed in reflection of the newest irregular events (zero interest rates, deflation, etc). Original model of CNB also doesn't use unemployment as an input, even though unemployment is used very often in the relevant literature, therefore this thesis shall investigate, whether unemployment should be a key variable in the credit risk satellite model. Competing new models are developed on basis of model combination methods (Bayesian Model Averaging, Frequentist Model Averaging) and Neural Networks. Models are developed on three different segments of loans - Coporate, Housing and Consumer loans. Comparison of the re-estimated CNB model with alternative models will be done on the full sample and then on 3 testing samples by training the models on the sample prior the testing period and then producing 8 quarters of predictions and comparing them to the real values.

The thesis is structured as follows: Chapter 2 overviews relevant literature on the topic from both Czech economy and from abroad. Chapter 3 describes Data used in the model and Methodology of the techniques used in developing forecasting models. Chapter 4 provides estimation results and comparison of results of competing models, consecutively for three segments of loans. Chapter 5 summarizes findings of the thesis and concludes the main findings.

Chapter 2

Literature review

Following chapter is focusing on overviewing relevant literature together with evolution of stress tests. This literature motivates the whole work of this thesis.

On the topic of stress testing and satellite models in particular there are not many studies published, even less studies in Czech context. The most relevant literature is last Czech National Banks published paper related to satellite models “Dynamic Stress Testing: The Framework for Testing Banking Sector Resilience Used by the Czech National Bank” by Gersl *et al.* (2012) which presents together whole stress testing practice used in Czech republic since the beginning. It briefly presents evolution of framework (which will be discussed further in this section later on), the whole procedure of stress testing from macro data through DSGE models through Satellite models to final predictions of financial variables which are then used using simple calculations to evaluate resilience of the banking sector. The most important part of this paper is presenting current methodology of constructing Satellite models by CNB. This framework in Czech republic didn’t change ever since the paper was published. These models are being re-estimated every period yielding new coefficients, which are used to make new predictions, however the models are remaining unchanged for many years, which is the main motivation for this thesis. Coefficients of every re-estimation are not published (with exception of this particular paper, where they are shown for one period), however predictions of those models are published, which will serve as a tool for comparison of models by CNB and different models developed in this thesis. Following this study from 2012, there are no papers published on this topic related to stress testing in Czech republic.

Second most relevant source and also the most recent study is “Stamp€:

Stress-Test Analytics for Macroprudential Purposes in the euro area" by Dees *et al.* (2017). This paper published by ECB, like the paper from CNB, provides current stress testing framework, this time for euro area. It is more or less conducted in similar spirit as Gersl *et al.* (2012). Main usage of this paper is that it provides current methodology by ECB that will be used (among other methods) to construct Satellite models for Czech economy.

Second paper to mention published by ECB is "A macro stress testing framework for assessing systemic risks in the banking sector" by Henry & Kok (2013) published earlier than Stamp€:. This study is more focused on theoretical practices of whole stress testing in a frame, describing the whole process through 4 pillars structure of solvency analysis. Scenario pillar where shocks are introduced, satellite model pillar where macroeconomic changes are translated to changes in financial variables. Third is balance sheet pillar which performs solvency position of banks using scenarios that came out from satellite models. Fourth feedback pillar shows the effects of solvency changes projected to the economy.

One of the earliest papers published on topic of stress testing is study from IMF by Blaschke *et al.* (2001). This study argues that most recent crises (with respect to year 2001), mosty financial crises in Asia, uncovered needs for developing more sophisticated approaches to financial risk management, since financial stability plays huge role in macroeconomic performance. Except putting together during those times current stress testing framework, in particular for credit risk he proposed one of the first macroeconomic variables to financial variables model - a.k.a. Satellite model. Blaschke proposed regression of ratio of NPL (non-performing loans) to Total assets on interest rate, inflation rate, change in real GDP and change in "terms of trade". However, the author noted that it was difficult to construct such models due to lack of data in modern economy that had it's modern form for only few years.

First paper from Czech Economy related to stress testing is "Effects of Macroeconomic Shocks to the Quality of the Aggregate Loan Portfolio" by Baboucek & Jancar (2005) published under CNB. Following similar literature mostly from US economy after year 2000, they tried to establish connection between macroeconomic variables and credit risk using Vector Autoregression (VAR) model. They realize that the strongest channel goes from macroeconomic variables to credit risk, however feedback response can be crucial as well, VAR model can tackle this concern. Similar to Blaschke, authors decided to use NPL as an indicator for credit risk as it seemed to be the most straight forward measure. They use VAR with 9 endogenous variables, besides NPL there is real exchange rate, exports, M2 money aggregate, imports, bank loans to clients, unemployment, CPI, domestic interest rate. Although signs of the results are mostly consistent with previous literature, only 8% of the coefficients were statistically significant, which can be most likely explained by short time series sample before year 2005.

Building on paper by Baboucek and Jancar, study from IMF by Cihak & Hermanek (2005) has been published. This paper states current conditions of Czech stress testing. It points out that Czech modeling lacks stronger 2 way links between stress tests and macroeconomic scenarios, in particular credit risk models.

Last relevant study from Czech economy is study by Jakubik & Schmieder (2008) about comparison of during those times new member state of EU Czech republic and the largest EU economy Germany. Study is related to time period between 1994 and 2006 consisting 3 stress periods — Asian and Russian crisis and financial crisis in 2001. Authors found that both economies can be modeled using similar macroeconomic variables, despite different default rate patterns in both countries. This study points out that after 2003 the economy moved towards lower risks as inefficient firms didn't survive changes around year 2000. The year 2003 importance is consistent with current CNB methodology (Gersl, 2012) where in their models they use data after year 2003 due to changes in economy. In the study default rate is used as a dependent variable. For Czech corporate credit model they find that default rate is mostly dependent on real exchange rate (they use 2nd lag), which is consistent with expectations since Czech economy is strongly export oriented. Other explanatory variables were lag of inflation, GDP and 4th lag of Credit-to-GDP ratio. Turned out that nor GDP growth rates, unemployment or interest rates don't contribute to prediction power since they are usually correlated with variables included in the model. The same didn't hold for Germany where nominal interest rate seemed to be the key explanatory variable, among GDP, production and Credit-to-GDP. For household credit model Czech and German model consist of different variables, for Czech economy the most important predictors were unemployment rate and real interest rate, for Germany it were household income and Credit-to-GDP ratio.

One of the biggest problems of stress testing came to the surface after financial crisis in 2008, where most of the stress tests predicted that banking sector should have remained stable. Unfortunately most of the above mentioned studies were already outdated during the crisis. Study by Papadopoulos *et al.* (2016) covers empirical exercise of satellite modeling for credit risk containing for us most important time period, the financial crisis of 2008 and sovereign debt crisis of 2010, using period of time from 2006 until 2013 with macro data for 15 EU countries. This study tries combination of different models. Candidate models are chosen from different permutations of models with one to nine explanatory macro variables, which results to 511 models. Through different techniques, from these models 22 meaningful one are chosen. Most frequent variables among these 22 surviving models appeared government debt, GDP and unemployment and national disposable income (NDI). From these, models with included NDI perform unexpectedly well, models with GDP or unemployment on the other hand perform poorly compared to those without these variables. According to authors it may suggest, why some older models could perform poorly, because GDP and unemployment, as the easiest to interpret, were forced to be included into models, whereas models without them, which were not considered, can perform better. Combinations were chosen with different kind of weights — sophisticated Bates-Granger weights (based on out-of-sample RMSE), equal weights (simple model averaging) and median prediction. Combinations were constructed on full 22 model space and trimmed space of 10 best (5 respectively) models by performance. Prediction power was then compared among baseline models and 3 different weighting types among 3 different types of trimming, based on six different performance measures. It turned out that any combination approach always performed better than any baseline non-combined model. Next interesting finding was that less complicated weighting scheme usually outperformed complicated one. Last interesting finding is that combinations of models from smaller model space usually outperformed less trimmed spaces. This paper therefore suggest to use model combination, which is not present in CNB framework.

In Germany, Macro stress tests: Technical Documentation from Deutsche Bundesbank (2015), satellite models for credit risk has two different models, for small banks and for large banks. Panel data model for both size of banks is estimated with fixed effect using Generalized Method of Moments (GMM) framework. Dependent variable is Net loan loss provision, which is expense to set aside to serve as an allowance for bad loans. As a independent variables they use lagged value of Net loan loss provision to total assets ratio, lagged GDP growth, Book equity to total assets ratio for each individual bank and RWA to total assets (Risk weighted ratio). Results are seasonally adjusted.

Chapter 3

Methodology

In this chapter I provide methodology that is currently used in CNB and other methods that will be used for estimating competing models to CNB model. First of them will be methodology of ECB, Bayesian Model Averaging. Next I will discuss methodology about other model combining methods. Last but not least, I will introduce Artificial Neural Networks Methodology, which is the most innovative approach about Satellite models.

3.1 Data

This section will provide information about sources of data and information about variables chosen for the analysis.
Dependent variable for this thesis is 3 Month Probability of Default (for 3 different segments separately). It is a probability that subject in a given cathegory defaults within next 3 months. These probabilities of defaults are taken from every years official CNB reports - from CNB Financial Stability Reports (CNB Financial Stability Report 2006, CNB Financial Stability Report 2007, CNB Financial Stability Report 2008/2009, CNB Financial Stability Report 2009/2010, CNB Financial Stability Report 2010/2011, CNB Financial Stability Report 2011/2012, CNB Financial Stability Report 2012/2013, CNB Financial Stability Report 2013/2014, CNB Financial Stability Report 2014/2015, CNB Financial Stability Report 2015/2016, CNB Financial Stability Report 2016/2017, CNB Financial Stability Report 2017/2018). From these reports, data for quarterly measured Corporate Probability of Defaults from Q1 of 2003 to Q3 of 2017 were collected, which is 59 observations. Probability of Defaults for Consumer and Housing segment are available for shorter period of time,

from Q3 of 2007 to Q3 of 2017 which is 41 observations.
Explanatory variables contain Pribor, GDP, CZK/EUR exchange rate, Government Debt to GDP ratio, Household Consumption, Consumption to GDP ratio, Compensation of Employees, Inflation, Unemployment and Property Price. All explanatory variables are collected with at least the same length as PDs, all variables are measured quarterly.
Pribor is Prague Interbank Offered Rate, which is estimate of rate, for which reference bank is willing to provide deposit to another bank on interbank market, 3 Month Pribor was collected from CNB ARAD public database. GDP is collected also from CNB ARAD database and it is captured in real term in milions of CZK. Last variable taken from CNB ARAD is CZK/EUR nominal exchange rate, it is measured as an average of exchange rate for the given quarter. Government Debt to GDP ratio is collected from Eurostat, it is expressed in milions of CZK in % of GDP. Household Consumption (and Household Consumption to GDP) is also collected from Eurostat and it is Final consumption expenditure of households and non-profit institutions serving households measured in milions of CZK. Compensation of Employees expressed in millions of CZK is total sum of wages, salaries and employer's social contributions. Inflation calculated by CPI is taken from OECD database and it's quarterly measured annual rate. From OECD are taken also Unemployment rates. Property Price is the real cost of housing. Data source is Eurostat database combined with the CNB Financial Stability Report 2009/2010. Table 3.1 shows summary statistics of the Data — Number of observations, minimum, maximum and standard deviation.

3.2 Methodology of CNB

Current CNB model, as described in Gersl (2012) is motivated by two-step approach. First step is General-to-specific model-selection (Gets) algorithm, which identifies a subset of potential explanatory variables. Second step is a selection of model among candidate models with all combinations of explanatory variables obtained from Gets with pre-specified number of lags, final model is chosen by out-of-sample RMSE. Gets algorithm allows for multiple approaches including Arimax, ARDL, Arfima or Setar type of models. Current credit risk for all three segments (Corporate, Consumer, Housing) is done by Arimax (Autoregressive Integrated Moving Average with exogenous variables).

As a first step in this thesis, the whole algorithm will be replicated with

Table 3.1:
Data — Summary Statistics

Variable	Obs	Min	Max	SD
3 Month Corporate PD	59	0.20	1.37	0.28
3 Month Household PD	41	0.38	1.59	0.22
3 Month Consumer PD	41	1.135	2.31	0.29
3 Month Pribor	59	0.28	4.21	1.15
Real GDP (Quarter)	59	728945	1173341	100813
CZK/EUR exchange rate	59	23.9	32.8	2.3
Debt/GDP	59	26.8	45.5	6.4
Household Consumption	59	328557	606331	67918
Compensation of Employees	59	257101	518530	63060
Inflation	59	-0.39	7.44	1.66
Unemployment	59	2.76	8.43	1.51
Property Price	59	78.2	117.4	10.2
Consumption/GDP	59	45.6	50.9	1.33

new data using all above mentioned basic time series estimation frameworks. Predictions from other, more advanced methods, will be then compared to CNB methodology.

3.3 Bayesian Model Averaging

First to compete with the current CNB methodology is Bayesian Model Averaging (BMA) currently used in ECB modeling. The satellite model is developed and explained in Henry (2013). As author states, it is particulary useful approach in satellite modelling with respect to relatively short time series, because it allows to combine models with less size together, which can use predictive power of many predictors. Model combination approaches are motivated by study of Moral-Benito (2015).

In first step, set of candidate equation is chosen. In ECB context Autoregressive Distributed Lag (ADL) is the framework. In general form ADL equation is

$$Y_t = \alpha + p_1 Y_{t-1} + ... + p_p Y_{t-p} + \sum_{k=1}^{k_i} (\beta_0^k X_t^k + ... + \beta_{q^k}^k X_{t-q^k}^k) + \epsilon_t \tag{3.1}$$

where p number of autoregressive lags. If set of predictor variables is denoted as K, this equation is estimated for every combination k_i of predictors from K and for every lag structure of independent variables up to a lag q^k for predictor k from 1 to k_i. Optimal lag structure is then chosen from candidates among one set of predictors by minimizing Akaike or Schwarz criterions. The maximum number of predictors in a single equation is set to predefined number K (with respect to length of time series). From this procedure we obtain candidate models with optimal lag structure for every combination of independent variables. The individual posterior coefficient is calculated as

$$h(\beta \mid y) = \sum_{i=1}^{I} P(M_i \mid y) \frac{f(y \mid \beta)h(\beta \mid M_i)}{f(y \mid M_i)} \tag{3.2}$$

where $f(y \mid \beta)$ is density function of dependent variable conditional on β, similarly $f(y \mid M_i)$ is density of dependent variable conditional on the model and $h(\beta \mid M_i)$ is density function of β conditional on the model. These fractions, one for each candidate model, are then weighted to obtain final coefficients. Weights $P(M_i \mid y)$ in this setup is chosen by performance of individual candidate models. They are set proportional to in-sample Bayesian Information as they in Sala-I-Martin *et al.* (2004).
Model Priors shall be in all models set to Uniform priors, therefore model prior is always set to $1/2^K$ where K is the total number of models. Coefficient density priors are not taken as Zellner's g priors, which are widely used. Instead of that, BIC approximation shall be used. For coefficient β_1 the posterior probability that the coefficient is included in the model is

$$Pr(\beta_1 \neq 0 \mid D) = \sum_{j:\beta_1 \in M_j} p(M_j \mid D) \tag{3.3}$$

such that $p(M_j \mid D) = \frac{exp(-BIC_j/2)}{\sum_{i=1}^{K} exp(-BIC_i/2)}$

BMA as current satellite modelling framework will serve as main competing model to CNB Arimax model.

3.4 Other model averaging techniques

3.4.1 Forecast combination

Another discussed model is Forecasting combination used in Papadopoulos (2016). This approach is sometimes considered as predecessor of FMA, which will be discussed later on. Both approaches in contrast to BMA don't require setting any priors, thus the estimators are determined only by data. Setting priors may be problematic generally already in setting them, or while dealing with conflicts between them, as pointed out in Hjort & Claeskens (2003).

In Forecasting combinations, from chosen set of predictors, all permutations of predictor variables will be modeled and used as candidate models, disregarding those with insignificant coefficients. Suggested by Papadopoulos, Surviving models will be then ranked according RMSE. As in Papadopoulos, combination of surviving models will be done according 3 different weighting schemes. First weighting will be with equal weights, in other words, final forecasts will be only simple average of forecasts among surviving models. Second weighting scheme will be according median weighting, which means setting final forecast equal to median of forecasts among surviving models. Last weighting scheme will give weights to individual forecasts proportionally to RMSE of predictions from surviving models. This procedure will be done on full set of candidate models as well as on different depth of trimming worst performing models. This procedure will result into 3 models (forecasts) per depth of trimming plus no trimming.

3.4.2 Frequentist Model Averaging

Similar approach to Forecast combining is Frequentist model averaging (FMA). In fact, under linear models, weighting of forecasts is equal to coefficients weighting, making these two approaches almost identical. The whole procedure of FMA is extensively explained for example in Wang *et al.* (2009).

For linear model the regression would be

$$y = X\beta + Z\gamma + \epsilon \tag{3.4}$$

where X is matrix of regressors that must be included in the model, Z is matrix of regressors that may or may not be included in the regression. The goal is to estimate β. If S is one candidate model, the estimate of β for model combination is

$$\hat{\beta} = \sum_{S} \lambda(S \mid data)\hat{\beta}_S \tag{3.5}$$

In standard non-combination approach the weight $\lambda(S \mid data)$ is set to 1 for our pre-chosen model and zero for all other models. In FMA, weights are chosen in a way that they can smoothly capture effect of more competing models, not assuming that the true "correct" model is the chosen one, rather allowing for more candidates and assume that the "correct" model is among them.

Of course, the crucial question is how to set the weights. One such approach, similar to one used in Forecast combining, is weights based on Information Criterion, which is the approach used in this thesis. In this setup however weights are not simply linearly assigned by information criterion. Buckland *et al.* (1997) proposed following weighting

$$\lambda_k = \frac{exp(-I_k/2)}{\sum_{u=1}^{K} exp(-I_i/2)} \quad , \quad k = 1, 2, ..., K \tag{3.6}$$

where I_k is AIC from k-th candidate model. This weighting is widely used ever since.

3.4.3 Neural Networks

Last model type used in this thesis are Neural Networks. Neural Networks in general try to simulate functionality of a brain by introducing in neurons that learn from each other. In simple linear model the output variable is simply weighted average of K predictors:

$$Y = \sum_{j=1}^{N} \beta_j X_j \tag{3.7}$$

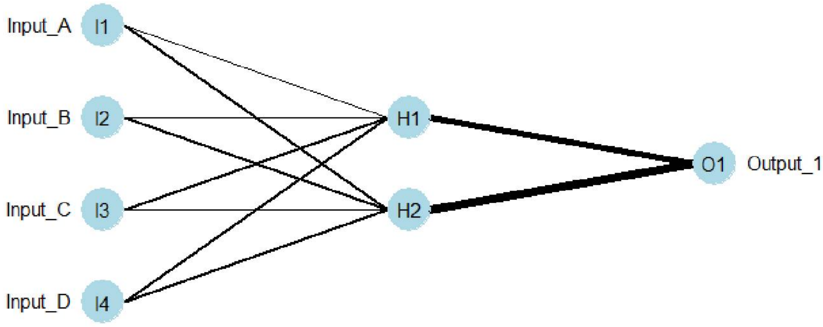
where β_j are weights. Neural Network (with one layer) on top of that introduce hidden neurons, which are themselves functions of the inputs. Neurons themselves then serve as inputs for the next step, where the final output (or outputs) is a function of hidden neurons. Hidden neurons are distributed in hidden layers. All hidden neurons in the 1st layer are functions of inputs, but they are not affected by other neurons in the same layer. Hidden neurons in the second layer are functions of neurons in the first layer, etc. Outputs are functions of neurons from the last layer.

Design with 1 hidden layer and 2 hidden neurons is shown in Figure 3.1. Outputs or hidden neurons don't have to be only linear combination of inputs. In general, output of Neural Network with one hidden layer is:

$$Y = h(\sum_{k=1}^{K} \alpha_k . g(\sum_{j=1}^{N} \beta_j X_j)) \tag{3.8}$$

where β are coefficients on inputs, α are coefficients on hidden neurons, h and g are so called activation functions.

Figure 3.1:
Neural Network - Basic design



Of course the crucial question is how to set the number of hidden layers and hidden neurons. Hidden layers will be always set to one, because more hidden layers would lead to overfitting with respect to short data sample. Number of hidden neurons within one layer are going to be number of predictors + number of outputs (which is in this case always one) divided by two. Activation function is going to be linear function. Procedure to set weights is going to be Standard Back Propagation algorithm, which is described in Rumelhart *et al.* (1986).

3.4.4 Comparison of model performances and predictions

For the purpose of comparison of estimated models, as a basic tool standard measurements will be used, such as Root Means Square Error (RMSE). RMSE of two vectors y and x is defined as

$$RMSE(y,x) = \sqrt{\frac{\sum_{u=1}^{n}(y_i - x_i)^2}{n}} \tag{3.9}$$

Second measure used is Mean Absolute Error (MAE), averages of MAE over are discussed for example in Willmott & Matsuura (2005). For two vectors y and x it is defined as

$$MAE(y,x) = \frac{\sum_{u=1}^{n} \mid y_i - x_i \mid}{n} \tag{3.10}$$

These measures however provide only theoretical performance comparison. They don't provide us any information about how each model would perform in situation for which it is developed in a first place — for stress scenarios. In addition to theoretical measures, three periods of time will be used to demonstrate prediction performances of all models. First one of them will be financial crisis period after year 2008. Models will be calibrated on available data prior year 2008. After that, from each model predictions for the recession period will be constructed. Construction of these predictions will be based on one crucial assumption and that is perfect foresight of all variables except financial variable of interest - probability of default. This assumption comes from the target of this thesis. Purpose of this work is not to question all prediction models of CNB, but only the satellite models. In this light instead of official predictions for the period of CNB for other variables than probabilities of default, actual historical values will be used. Predictions will be then compared among each other. Predictions will be compared graphically to clearly demonstrate, whether this thesis is successful in developing not only better models during non-stress periods, but as well if it tackled one of the main concerns about current CNB methodology — the fact that current models is not sensitive to shocks. Second period for which models will be compared will be period of years 2012 and 2013, which is the time economy recovery and mainly decline of the Probability of Default. Last period for which models will be compared in similar manner will be window of newest available data, years 2015 and 2016.

Chapter 4

Results

This section provides procedure and results for estimating models of credit risk for each framework stated in the methodology section as well as a comparison of performance with each other. Models are constructed for three different segments of loans — Corporate, Housing and Consumer.

4.1 Corporate loans

This section discusses prediction of Probability of Default for Corporate loans sector. Dependent variable is a level of 3M Probability of Default in Corporate sector measured quarterly. Data available are for quarters Q1 of 2003 to Q3 of 2017, which results into 59 data points. All models used in this thesis are allowing for up to the 4th lag of PD, which effectively cuts out 4 data points from the time series, resulting into 55 points in total.

4.1.1 CNB Methodology Framework

Current Methodology for credit risk model of corporate PDs estimated by Arimax model has these seven independent variables: 4th lag of Probability of default, 3M Pribor and its 1st and 2nd lag, de-trended value of the first difference of CZK/EUR exchange rate and its second lag and 4th lag of real GDP YoY growth.

Full Sample

Table 4.1. shows coefficients and standard errors from the simple Arimax regression on the full sample. The first column shows coefficients published in

Table 4.1:
Corporate Probability of Default

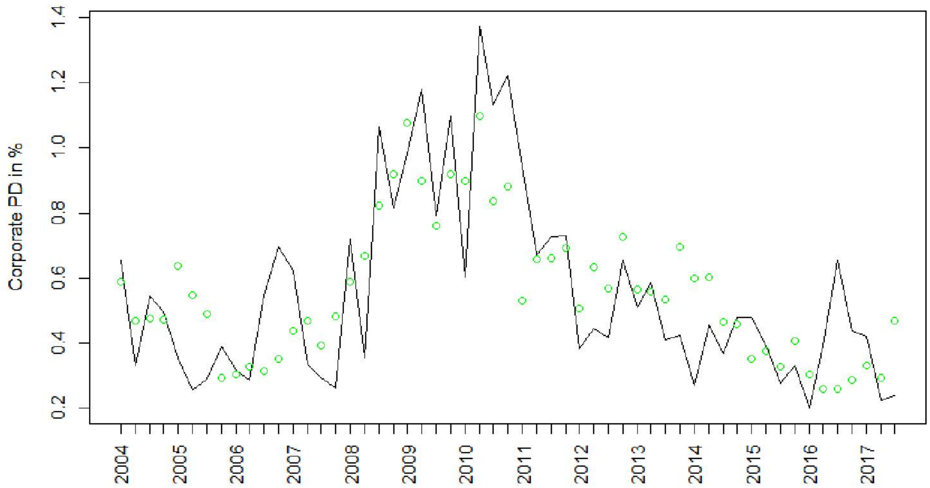
Dependent variable 3M PD_t	(1) CNB model 2012	(2) Re-estimated model
Constant	1.332 ***	0.173
	(0.155)	(0.160)
3M PD_{t-4}	-0.179	0.210
	(0.125)	(0.114)
3M $Pribor_t$	0.014	-0.090
	(0.073)	(0.105)
3M $Pribor_{t-1}$	0.057	0.030
	(0.082)	(0.161)
3M $Pribor_{t-2}$	-0.177 *	0.230 *
	(0.083)	(0.103)
$\Delta CZK/EUR_t$	-0,031	0.062
	(0.087)	(0.035)
$\Delta CZK/EUR_{t-2}$	0.085	0.024 *
	(0.071)	(0.037)
GDP	-0.074 ***	-0.039 ***
YoY $growth_{t-4}$	(0.016)	(0.011)
N	30	55
Adjusted R2	0.435	0.476

This table shows comparison of coefficients of Corporate PD model between original model from Gersl (2012) and re-estimated model for purpose of this thesis. Standard Errors in parenthesis. Significance signs: *5%, **1%, ***0.1%

Gersl (2012), second column shows results of re-estimated model on data up to Q3 of 2017.

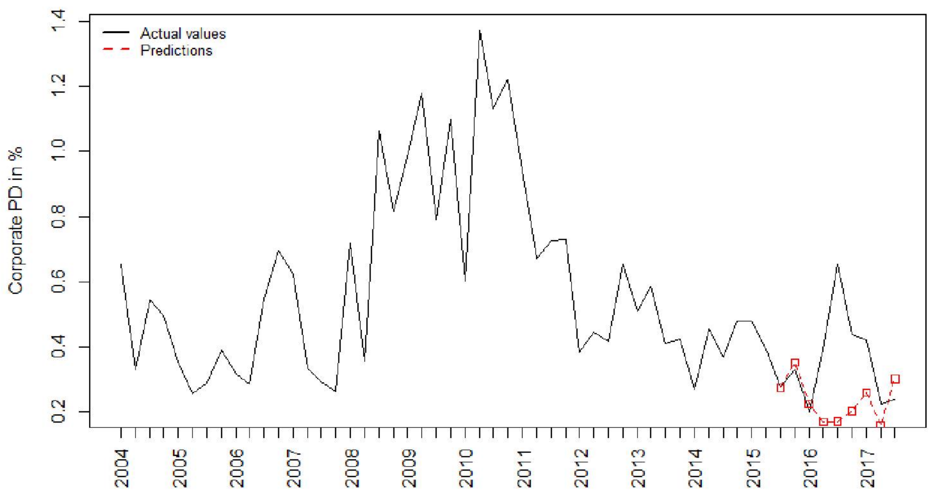
From this comparison of coefficients we can see that results are slightly different. Results are comparable in case of the 4th lag of GDP growth in sign and significancy. Coefficient of 2nd lag of Pribor has different sign, regaining significance on the other side of zero. Concerning 2nd lag of difference in CZK/EUR exchange rate, new results became significant on 5% level. Other variables retained their insignificancy on 5% level, however 4th lag of Probability of Default and current value of exchange rate difference are now significant on 10% level. These differences suggest that current methodology of CNB is very likely to be unstable in time. Fitted values of the estimation on the full data set are in Figure 4.1. RMSE of the fitted values versus Actual values is 0.1898, MAE 0.1560. These values will serve as a baseline for comparison of other methods used on the full data sample.

Figure 4.1:
CNB Framework - Fitted values



This graph shows fitted values of the original Arimax model (green circles) re-estimated on the full sample versus the data (solid line) for Corporate segment

Figure 4.2:
CNB Framework - Testing period 2016 predictions



Testing period 2016

The same Arimax regression is used for testing the out of sample performance during the first testing period, which is 8 quarters between Q4 of 2015 and Q3 of 2017. Model is re-calibrated on shorter training period consisting of 47 periods up until Q3 of 2015. Predictions for the testing period are constructed in a way that data for all variables (and their lags) are taken as actual values. Lags of PD are considered in a way that values from the training period are taken as the actual values, predictions themselves serve as values for the testing period. In this way, 8 predictions are constructed and shown in Figure 4.2. Predictions are visualised in a way that the first point is the last known point, next 8 points are the predictions themselves.
Predictions are very close to the actual values in the first two points, capturing increase of PD and then drop. However following spike during Q2 of 2016 to Q2 of 2017 isn't captured at all. In contrary, the model predicts further decrease. RMSE of Actual values versus 8 predicted points is 0.2160, MAE is 0.1590.

Testing period 2012

Next testing period of 8 quarters is period Q1 of 2012 till Q4 of 2013. Predictions are constructed in the same way as in the previous case. Training sample is in this case 32 data points. Last known value and next 8 predictions are in Figure 4.3.
Predictions this time didn't capture decline of the high values in previous years and predicted return of PDs to similar values as in the crisis periods. RMSE of these 8 predictions versus Actual values is 0.6246, MAE is 0.6039.

Testing period 2008

Last testing period of 8 quarters is period of Q3 of 2008 to Q2 of 2010. Training sample have only 18 data points, so the estimation suffers from a small amount of data points. Last known value and 8 following predictions are in Figure 4.4. Even though it may seem like the model did a good job, since some of the predictions are very close to the Actual values, including the highest peak, we can see that predictions are very volatile, predictions even consist one negative value. This is a consequence of the short training period. RMSE is 0.4990, MAE 0.3336.

Figure 4.3:
CNB Framework - Testing period 2012 predictions

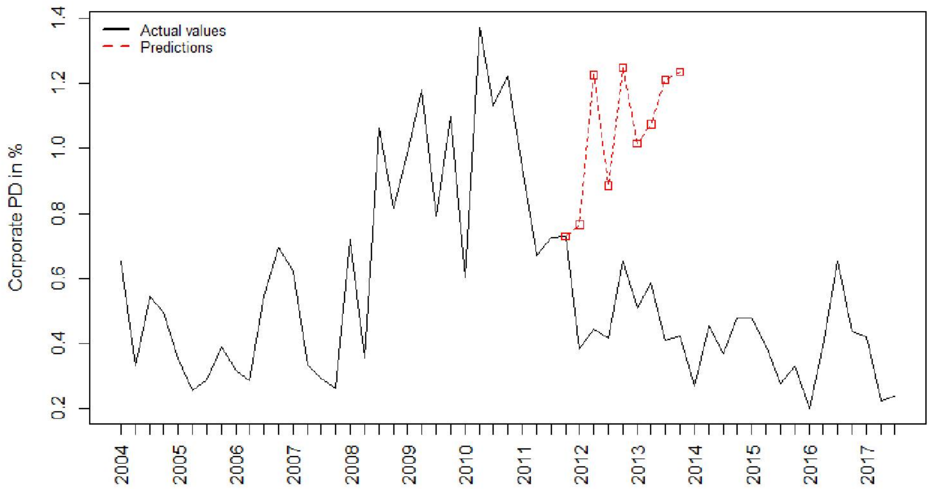
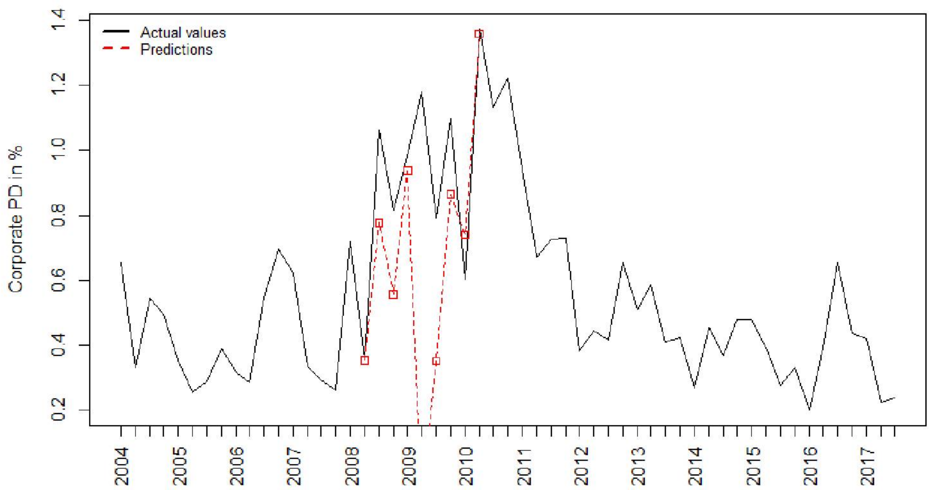


Figure 4.4:
CNB Framework - Testing period 2008 predictions



4.1.2 Forecast combination

Approach from Papadopoulos (2016) will be used as the first preliminary analysis concerning Forecast combination approach. He demonstrates advantage of Forecast combination on simple linear modeling with independent variables from period t with no lagged values. He shows that surviving forecast combinations under many different settings always perform better (according to common MAE and RMSE measures) than any non-combined linear model estimated from the subsets of the same set of variables.
We have chosen 10 independent variables as follows: Pribor, GDP in levels, Exchange rate in levels, Debt to GDP ratio, HH Consumption, Compensation of employees, Inflation, Unemployment, GDP growth and Consumption to GDP. The strategy is to estimate linear model with dependent variable Probability of Default of corporate sector with all possible combinations of subsets from these 10 independent variables, always with time period variable included to deal with possible time trend. This results into $2^{10} = 1024$ linear models. Next step is to keep only those models, where all independent variables are statistically significant on 5% level. This leaves us with 32 surviving models. The most frequent independent variable among surviving models is Household Consumption, it appears in 16 out of 32 surviving models, 2nd and 3rd behind are CZK/EUR exchange rate and Consumption to GDP ratio, they appear 13 and 12 times respectively. The least frequent variable is Debt to GDP ratio, which appears 3 times. All other variables are present in 5 to 10 out of 32 models.
For all 32 models we obtain predictions. To combine these prediction we will use 3 weighting schemes: average of predictions, median predictions and Bates-Granger weights. Weight of every model in average weighting is simply

$$w_i = \frac{1}{m},\ i \leq m \tag{4.1}$$

where i is number of corresponding candidate model and m is the total number of surviving models.
In median weighting the weights are

$$w_{i,t} = \begin{cases} 1, & \hat{y}_{i,t} = median_{j=1,\ldots,m}(\hat{y}_{j,t}) \\ 0 & \end{cases} \tag{4.2}$$

which means that for each time period, the final combined prediction is median of predictions among candidate models within the same time period.

Last weighting scheme is Bates Granger weights, where weights are distributed as following

$$w_i = \frac{\hat{\sigma}_i^{-2}}{\sum_j^m \hat{\sigma}_j^{-2}} \tag{4.3}$$

where $\hat{\sigma}_j$ is RMSE of j-th candidate model.
In case of full surviving models space, this procedure is made on all 32 models. Performance can be however improved by trimming out some worse performing models out of full space of candidate models. For this purpose, after making predictions for each individual model, all models are sorted by RMSE from the best to the worst performing one. Then certain portion of worst performing individual models are disregarded. Among models that has been kept, weights are again assigned by the same rules as above.
Finally, we have 32 individual prediction vectors, 3 combined prediction vectors from full space of competing models and 3 more prediction vectors for each depth of trimming.
Now we can investigate the performance of each version. Out of 32 candidate models, the minimal RMSE is 0.1854 for model with independent variable GDP, CZK/EUR exchange rate, Inflation and Household consumption. Maximum RMSE from individual models is 0.2669, average is 0.2268. If we take a look on combined predictions from full space of models, under average weighting we obtain RMSE of 0.2087, 0.2057 under median predictions, under Bates-Granger weights 0.2059.
These results go against result originally obtained in Papadopoulos (2016), where author finds out that all combination schemes perform better than every individual linear model. These results don't support that claim. RMSE under Bates-Granger weights is higher than in case of 5 best individual models out of 32 original. In case of average weighting, it is worse than 7 best models. This as well goes against the findings in Papadopoulos, where they find out that less sophisticated weighting schemes usually obtain better results.
If we however consider trimmed space of candidate models, we obtain better results. After trimming bottom half of candidate models, from average weighting we obtain RMSE of 0.1985, from median predictions 0.1978. With Bates-Granger weights among 45 models we improve RMSE to 0.1976. We still did not achieve lower RMSE than in case of best linear individual model, but under Bates-Granger weights as well as under average weighting, only 2 individual models reach to lower RMSE.

Table 4.2:
Corporate PD - Forecast combination

Full space	RMSE	MAE
Best individual model	0.1854	0.1458
Bates-Granger weights	0.2059	0.1634
Median predictions	0.2057	0.1640
Mean predictions	0.2087	0.1652

50% best models	RMSE	MAE
Best individual model	0.1854	0.1458
Bates-Granger weights	0.1976	0.1571
Median predictions	0.1978	0.1587
Mean predictions	0.1984	0.1577

25% best models	RMSE	MAE
Best individual model	0.1854	0.1458
Bates-Granger weights	0.1930	0.1525
Median predictions	0.1943	0.1555
Mean predictions	0.1934	0.1529

This table shows results of Forecasting combination approach, preliminary analysis of model combination approaches.
Table shows comparison of the best individual model versus three different weighting techniques for three different depths of trimming the model space.

If we trim the space while keeping only 8 best individual models, we already obtain RMSE lower than second best individual model under all weighting schemes. Any weighting scheme under any depth of trimming doesn't achieve lower RMSE than individual model with the lowest RMSE.
Second measure that we investigate is MAE. If we construct Bates-Granger weights based on MAE and then trim space according to MAE instead of RMSE, we obtain different results. Best individual model according to MAE remain the same, however the order of the others is changed. MAE of the highest ranked model is 0.1459. Even this time the combination coming from the full space of models doesn't yield better result as well as trimming the individual models. Results are summarized in Table 4.2.

4.1.3 Frequentist Model Averaging

Following previous section, where the first analysis about model combination approach was done, Frequentist Model Averaging (FMA) will be the first model

to challenge the original Arimax model. Unlike in the case of forecast averaging, this time we already include lags in the modeling procedure. This however brings other problems, mainly the size of the candidate model space. It was necessary to estimate $2^{10} = 1024$ models without any lags in the last section. If we want to include 4 lags for each of the 10 independent variables plus 4 lags of Probability of Default, it would result into 2^{54} models, which is more than 10^{16} models. To reduce the candidate model space, it is necessary to cut out some variables and/or some lags. In order to keep in line with the original Arimax setup, it is necessary to consider lags of Probability of Default up to the 4th order. From independent variables, by the same logic, 3M Pribor up to the 2nd lag, GDP YoY growth up to the 4th lag and first difference of CZK/EUR exchange rate up to the 2nd lag. In addition, in a line with a literature and with the aim of this thesis, Unemployment and its 1st and 2nd lag are included as well. This leaves us with $2^{18} = 262144$ models.
Second option how to decrease the model space is to use the model selection procedure, which will be used later in BMA section (and is used in ECB methodology as well). This candidate models selection will allow us to consider more variables. 6 non-lagged predictiors are chosen: 3M Pribor, Consumption to GDP ratio, Inflation, Unemployment, GDP growth and the first differece of CZK/EUR exchange rate. These 6 predictors and their lagged values up to the 4th order together with lagged values of PD up to the 4th order and time trend variable form 35 potential predictors. For all possible combinations of 6 predictors, optimal lag structure of the ARDL regression is found by minimizing the Bayesian Information Criterion. This results into $2^6 = 64$ candidate equations. ECB methodology selection requires all the equations to be without gaps. It means that if k-th lag of any variable is not included in the individual regression, lag of order k-1 (and therefore k-2, k-3 etc.) can't be included as well. FMA (as well as BMA) will be estimated with this no gaps restriction as well as without the restriction.

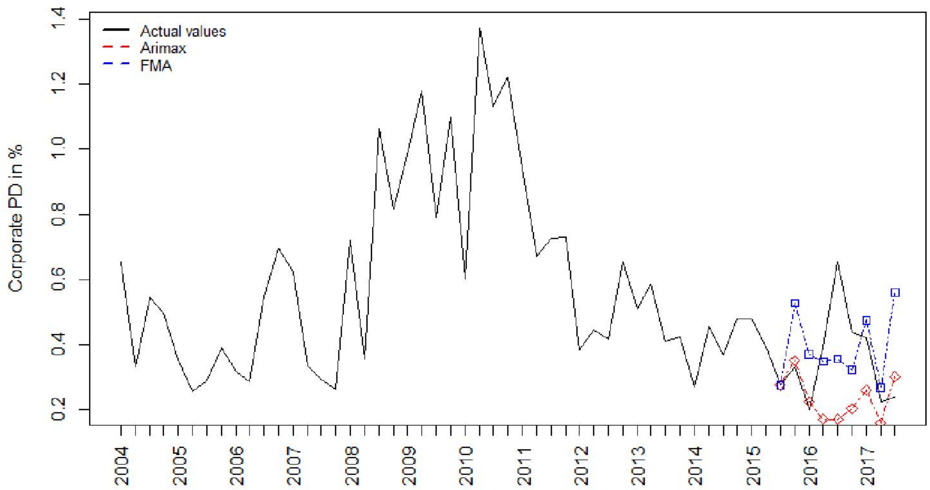
Full Sample

Similar procedure to the forecast averaging is performed in case of FMA. All 262144 candidate models are estimated (with included time period variable to deal with a possible trend) and then the model space is restricted only to models, which have all coefficients (except intercept and trend) significant. This results into 341 surviving models. For these models, weighting scheme proposed Buckland, Burham and Austin based on information criterion (described in the methodology section) is used to assign weights for surviving models. This weighting scheme is based on Akaike Information Criterion (AIC) and individual weight is a monotone decreasing function of AIC. The biggest weight (8.97%) is assigned to a model with independent variables being 1st lag of PD, GDP growth and its 3rd and 4th lag, 1st difference of CZK/EUR exchange rate and its 1st lag and 2nd lag of Unemployment. The most frequent variable among surviving models is 4th lag of GDP growth (in 184 out of 341 models), followed by 1st lag of PD (133 times), 1st difference of CZK/EUR rate (117 times) and 3rd lag of GDP growth (112 times). All other variables are present in 40 to 100 cases with exception of 2nd lag of 1st difference of CZK/EUR rate, which is present only in 6 cases. First column of Table 4.3 shows the weighted coefficients for all 18 variables from the restricted model space.

After obtaining coefficients from this setup of FMA model and after we construct predictions, we can compare the model to re-estimated original Arimax model. Before that, it is useful to check, whether the combined model is performing better than model with the largest weight. In previous case of forecast combination, it was not the case. The best model (according to AIC) from all candidate models has RMSE of 0.1603. The combined model however reaches to RMSE of 0.1569, therefore according to this measure, the combined model is better than the best model from the candidate model space. Now the combined model can be finally compared to Arimax re-estimated model. RMSE of Arimax model is 0.1898. Therefore FMA method was successful to obtain better performing model (according to RMSE) than the best individual model and as well better than the original Arimax model. The same conclusion is reached in case of the MAE comparison. Original CNB model has MAE of 0.1560, the best individual model 0.1253 and the combined model 0.1224.

If we don't restrict ourselves only to models with all variables being significant on 5% level and we rather keep all 262144 models, the best performing model from the restricted case has higher AIC (so it is performing worse according to this measure) than 5 other models with some coefficients being insignificant. Weight of the best performing model from the restricted case is now 0,08% and weight of the best model from unrestricted case is 0.13%. Set of variables included in the best several models in the unrestricted case is very similar to each other, more similar than in the case of restricted case. This suggests that in

Figure 4.5:
FMA - Testing period 2016 predictions



the unrestricted case, the best performing setup will have even stronger weight, since there are many models very similar to each other (different in only few, mostly insignificant coefficients).

If we compare performance of FMA based on the unrestricted model space (coefficients are in Table 4.3, second column), we obtain RMSE of 0.1548 (lower than the best performing model), which is lower than RMSE of FMA model from the restricted model space. Different result is case of MAE, it is 0.1233, which is higher than than in case of restricted model space. In both cases the FMA setup performs better than original Arimax model and the best performing individual models, however it is unclear from these measures, if restricting the model space improves the performance or not.

If model selection based on ECB methodology is considered, the best individual model with restriction of no gaps has RMSE of 0.1837, the combined model has 0.1813. MAE of the best individual model is 0.1479, combined model has 0.1412. Both measures are better than in case of Arimax model, however both are worse than in previous cases of FMA.

Relaxing the no gaps restriction completely provide us the best model model among the all the estimated FMA setups. The best individual model has RMSE of 0.1407, combined model 0.1339. MAE is 0.1163 and 0.1080 respectively.

Table 4.3:
FMA coefficients

Dependent variable PD_t	(1) restricted space	(2) full space
Constant	0.838	0.480
Period	-0.005	-0.001
3M Pribor_t	0.0008	-0.03
Unemployment	-0.010	-0.02
GDP YoY Growth_t	-0.021	-0.014
$\Delta CZK/EUR_t$	0.036	0.035
PD_{t-1}	0.285	0.254
PD_{t-2}	0.139	0.238
PD_{t-3}	0.0005	-0.008
PD_{t-4}	0.00008	-0.009
3M Pribor_{t-1}	0.028	0.055
3M Pribor_{t-2}	0.017	0.059
$\mathrm{Unemployment}_{t-1}$	0.018	0.060
$\mathrm{Unemployment}_{t-2}$	-0.058	-0.069
GDP YoY Growth_{t-1}	-0.002	0.002
GDP YoY Growth_{t-2}	-0.001	0.0008
GDP YoY Growth_{t-3}	0.042	0.027
GDP YoY Growth_{t-4}	-0.069	-0.056
$\Delta CZK/EUR_{t-1}$	-0.007	-0.054
$\Delta CZK/EUR_{t-2}$	0.0007	0.006

This table shows coefficients of two FMA models, which are not using the ECB candidate model selection procedure.
The first column stands for the model where individual models with all coefficients being significant on 5% level are considered. Second column model doesn't require this condition.

Testing period 2016

In the similar manner as in section investigating original Arimax regression, predictions for testing period from Q4 of 2015 till Q3 of 2017 are constructed. Regression are not only re-calibrated on the training sample of 48 points, but in case of selection of candidate models similar to ECB methodology, the selection is performed on the shorter training sample as well. The last known point as well as the 8 quarters of predictions from the best FMA framework from the full sample estimation (ECB model selection, no restrictions on gaps) are in Figure 4.5. The results are compared to the the Arimax Framework predictions for the same period.
On the graph we can see that the FMA framework follows similar movement directions as the Arimax model, however closer to the spike during 2016-2017. RMSE of these 8 points from FMA is 0.1868, which is lower than 0.2160 of Arimax regression. MAE of FMA is 0.1551, Arimax has 0.1590. According to the both measures, FMA performed better than the Arimax model.

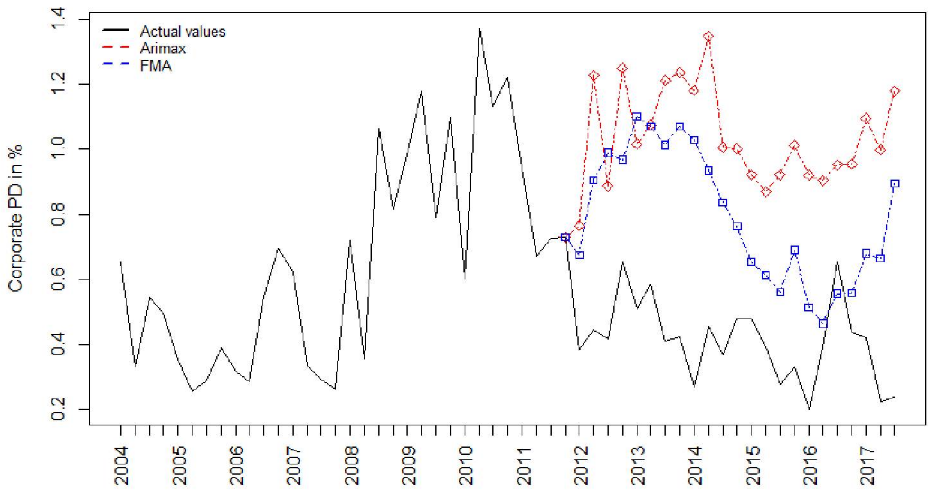
Testing period 2012

Figure 4.6. shows predictions of the best performing FMA model (ECB selection, gaps allowed) on testing period starting on Q1 of 2012. Predictions are again compared to the Arimax framework computed earlier. Graphs this time shows not only the last known point and next 8 quarters of predictions, but predictions until the very last data point as well.
FMA framework as well as Arimax model didn't capture recovery of PDs to low levels at first, even though FMA predictions are closer to the Actual values. Unlike the Arimax model, FMA predictions however predict decline of the PD later, after another predicted spike during years 2012 - 2014. RMSE of the first 8 FMA predictions is 0.5106 versus 0.6246 of Arimax model, MAE is 0.4950 versus 0.6039. Therefore even though the predictions of FMA didn't predict well the immediate decline of PD, it performed better than the original Arimax.

Testing period 2008

Last known point and next 8 predictions of the best performing FMA model compared to Arimax framework on testing period starting in Q3 2008 are in Figure 4.7.
FMA during this testing period suffers from the same troubles as the Arimax,

Figure 4.6:
FMA - Testing period 2012 predictions



being short training sample consisting of only 18 points. Predictions are very unstable, but still closer to the Actual values than Arimax framework. In fact, first 7 predictions are close to the Actual values, however the 8th last prediction suggests decline of the high PDs, unlike the 8th prediction of CNB Arimax, which almost hits the actual value. RMSE on these 8 testing periods is 0.5097 versus 0.4990 of Arimax. MAE is 0.3704 versus 0.3340. According to these measures, FMA didn't outperform Arimax model in this Testing period.

4.1.4 Bayesian Model Averaging

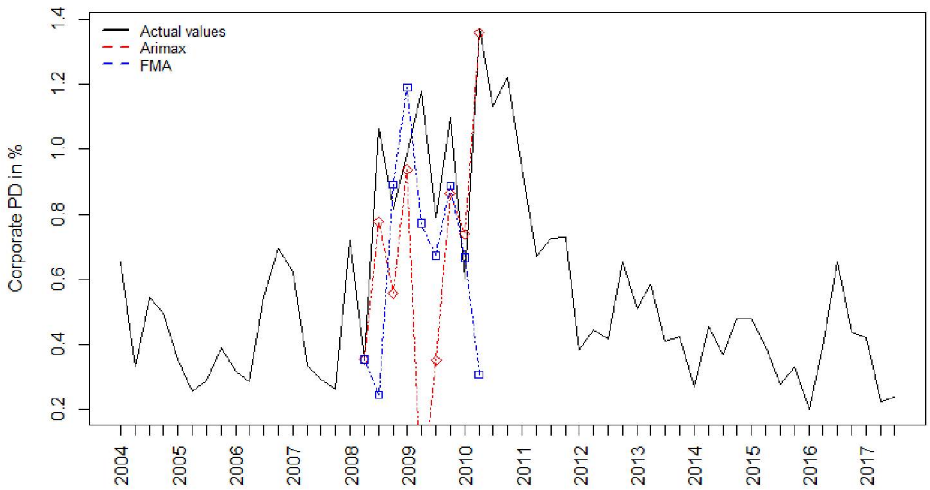
Main competing framework to the current CNB methodology shall be current framework of ECB. Dependent variable Probability of Default is in the ECB framework transformend by the logit transformation:

$$Y_t = log(y_t) - (1 - log(y_t)) \tag{4.4}$$

This ensures that the final predictions of the Probability of Default are bounded to the interval (0,1). Predictions from BMA are transformed back from logit version to final predictions by sigmoid funcion:

$$y_t = \frac{exp(Y_t)}{1 + exp(Y_t)} \tag{4.5}$$

Figure 4.7:
FMA - Testing period 2008 predictions



The same 6 predictors as in the FMA models according to ECB selection are chosen: 3M Pribor, Consumption to GDP ratio, Inflation, Unemployment, GDP growth, first difference of CZK/EUR exchange rate. These 6 predictors and their lagged values up to the 4th order together with lagged values of PD up to the 4th order and time trend variable form 35 potential predictors for the BMA procedure.

Full sample

In the same way as in FMA section, for all possible combinations of 6 predictors, optimal lag structure of the ARDL regression is found by minimizing the Bayesian Information Criterion. ECB methodology requires the equation to be without gaps. It means that if k-th lag of any variable is not included in the individual regression, lag of order k-1 (and therefore k-2, k-3 etc.) can't be included as well.

This results into $2^6 = 64$ candidate equations. These equations are not any further restricted, all of them remain in the model. The condition of no gaps turns out to be very restrictive in BMA as well as in FMA, cutting out most of the lagged values completely. Only the 1st lag of PD is included in all the candidate equations. The 2nd lag is included only in 12.5% of equations. Only 2 more lagged values are present in at least some equations, being the 1st lag

3M Pribor in 26.6% of equations and the 1st lag of CZK/EUR exchange rate in 10.9% of equations. All other lagged values are left out from all the equations due to the no gaps condition, because all the versions with other lags with no gaps are suboptimal. The best model from this setup contains only 2 variables, GDP growth and 1st lag of PD, this individual model accounts for posterior probability of 23.6%. All the best 5 models have only 2 or 3 predictors (+ time trend) and all account for cumulative posterior probability of 51.8%.

Since the no gaps condition turned out very restrictive, the same procedure is performed a setup with loosening the no gaps condition, this is justified for example by the CNB Arimax model, where gaps between lags are allowed as well. This setup allowed for the most of the potential regressors to be in at least some models. The best five models (according to the posterior probability of inclusion) consist of 3 to 8 independent variables (+ time trend) . Cumulative posterior probability of best five models account for 50.2%. Both setups however don't perform that well compared to previous combinations method. After transformation of the predictions by the sigmoid function back from logit transformed PDs, RMSE of the setup without gaps is 0.2092. Allowing for any lag structure, RMSE got below Arimax model, to 0.175. Recalling the RMSE of the original Arimax model being 0.1898. Considering MAE as the measure of the performance, both setups obtain MAE slightly better than Arimax setup, however it doesn't suggest these models to be performing that well, such as for example FMA. One big difference is however present between these BMA models and all previous setups. The logit transformation hasn't been used in any of them.

If we don't use the logit transformation and keep the PD (and lags of PD) as levels, similar results in terms of inclusion of lagged values are obtained. In no gaps setup, only three lagged values are used in at least one optimal equation, being it the 1st and 2nd lag of PD in 100% of the models and the 1st lag of 3M Pribor in 33.7%. When allowing gaps, 19 lagged variables make it at least one individual model, lags of PD are again no longer in all of the models. Number of predictors in the best models are very similar, so is the cumulative posterior probability. Posterior inclusion probabilities are summarized in Table 4.4, columns (1) and (2).

Not transforming PD into logit version improved performance slightly. Both models RMSE and MAE are better than Arimax model. Full power of the BMA setup will be however explored in the stress periods comparison.

We can see that loosening restrictions in this case always improved the in-sample performance. Therefore the last option in constructing BMA is in loosening not only no gaps condition, logit transformation, but it the procedure of choosing the candidate models as well. In other words, using all 35 potential predictors in the BMA model. This results into more than 30 billions of models, so the model space have to be restricted somehow anyway. This is done by the condition that only 150 best models of each model size (number of predictors included) are included are considered, based on BIC. This space is further restricted by Occam's window restriction (BIC of the individual model be bigger than minimum of bic among all models by more than $2.log(20)$). This procedure left us with 1116 candidate models. This model accounts for much higher uncertainty among the best models. Posterior probability of the best models in previous setups were around 20%, now it is less than 1% with cumulative posterior probability of best 5 models only 4%. We can see that this setup accounts for much bigger variety of predictors. These inclusion probabilities are partially inconsistent with the CNB model variable selection, since most of the variables have higher inclusion probability on different lags than used in the Arimax model. Posterior inclusion probabilities of this BMA model are in column (3) of Table 4.4.

Considering performance according to RMSE and MAE, this setup without restrictions from ECB setup yield significantly better results than the Arimax framework and previous BMA setups. RMSE is now 0.1455 versus 0.1898 of Arimax. MAE, is 0.1170 in case of this BMA versus 0.1560 of the Arimax model.

These results suggest that BMA setup is a very good candidate to outperform Arimax model.

Testing period 2016

The same testing periods as in FMA and Arimax setups are investigated in BMA setup as well. Predictions from the best performing model (on the full sample) during the period from Q4 of 2015 to Q3 of 2017 are shown in Figure 4.8. These 8 predictions are again compared to the original Arimax predictions from the same testing period.

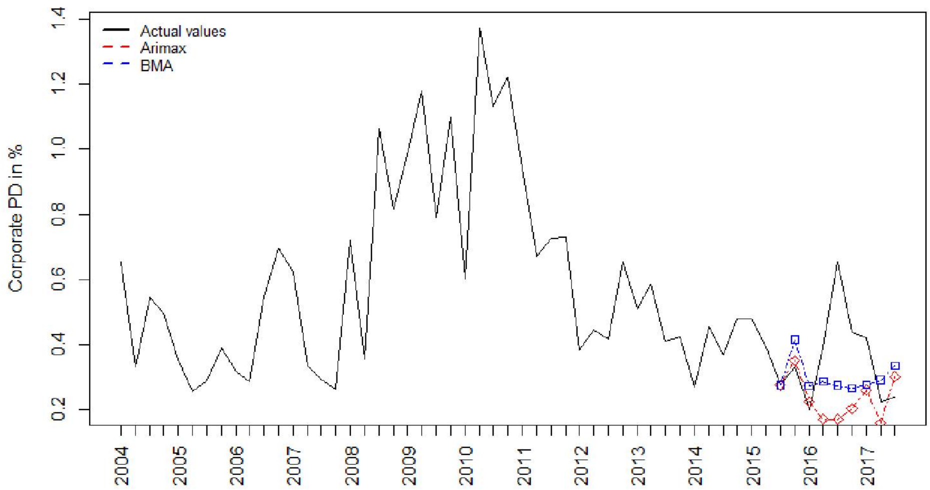
In similar way as FMA, BMA didn't capture the last spike during years 2016-2017. However predictions are closer to the Actual values. RMSE of these 8 predictions is 0.1703 versus 0.2160 of Arimax. MAE is 0.1401 versus 0.1590. Both measures show that BMA performed better than Arimax model.

Table 4.4:
BMA posterior inclusion probabilities

	(1)	(2)	(3)
Intercept	100.0	100.0	100.0
3M Pribor$_t$	39.6	2.5	20.0
Cons/GDP$_t$	14.7	0.5	8.4
Inflation$_t$	18.6	2.7	2.2
Unemployment$_t$	17.2	8.4	19.6
GDP YoY Growth$_t$	30.1	70.6	42.3
$\Delta CZK/EUR_t$	23.0	34.1	9.3
PD$_{t-1}$	100.0	19.2	13.7
PD$_{t-2}$	100.0	94.8	92.4
PD$_{t-3}$			10.6
PD$_{t-4}$			1.0
3M Pribor$_{t-1}$	33.7	8.6	16.5
3M Pribor$_{t-2}$			13.5
3M Pribor$_{t-3}$			10.5
3M Pribor$_{t-4}$		86.0	81.4
Cons/GDP$_{t-1}$			1.2
Cons/GDP$_{t-2}$		0.2	1.5
Cons/GDP$_{t-3}$		2.1	6.0
Cons/GDP$_{t-4}$		66.8	35.3
Inflation$_{t-1}$		0.5	6.3
Inflation$_{t-2}$		3.9	11.7
Inflation$_{t-3}$		14.8	10.2
Inflation$_{t-4}$		24.5	65.4
Unemployment$_{t-1}$		9.9	3.3
Unemployment$_{t-2}$		1.5	4.0
Unemploymentt-3		8.4	50.0
Unemploymentt-4		8.0	3.9
GDP YoY Growtht-1			3.2
GDP YoY Growtht-2			1.9
GDP YoY Growtht-3		26.6	20.0
GDP YoY Growtht-4		91.9	53.3
d CZK/EURt-1		11.2	19.2
d CZK/EURt-2			4.4
d CZK/EURt-3			1.9
d CZK/EURt-4		26.3	22.5

This table shows Posterior Inclusion Probabilities of individual predictors for BMA. Column (1) stands for model without logit transformation with ECB selection and no gaps restriction. Column (2) is model without logit transformation, ECB selection, gaps allowed. Column (3) model without logit transformation, no ECB selection

Figure 4.8:
BMA - Testing period 2016 predictions



Testing period 2012

Not restricted BMA model by ECB selection procedure again performed the best among other BMA setups during testing period 2012. Predictions compared to Arimax setup are shown in Figure 4.9.
We can see that unlike FMA and unlike Arimax, this BMA setup was successful in capturing the PD decline. RMSE of these 8 predictions is 0.2330 versus 0.6246 of Arimax predictions, MAE is 0.2200 versus 0.6039. BMA in this testing period outperformed Arimax model by far.

Testing period 2008

The best performing BMA model in Testing period 2008 is this time the model with ECB selection procedure, without restrictions on the no gaps condition. Last known point and 8 predictions compared to Arimax prediction are in Figure 4.10.
Comparison of this BMA setup and Arimax is unclear. BMA doesn't have that unstable predictions as both Arimax and FMA have in this testing period, however it doesn't capture the highest peaks very well. RMSE of these 8 predictions is 0.4505 versus 0.4990, MAE is 0.3616 versus 0.3336. According to these results we can't say that BMA outperforms Arimax model during this testing period.

Figure 4.9:
BMA - Testing period 2012 predictions

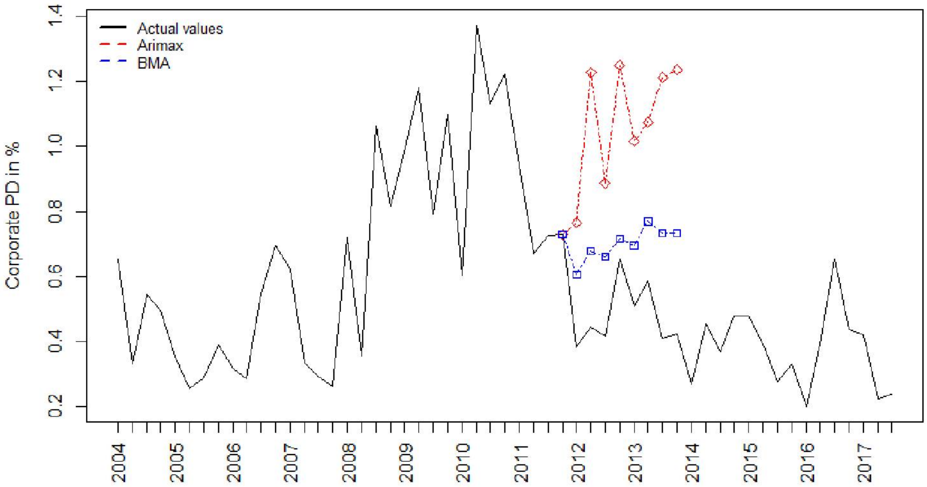
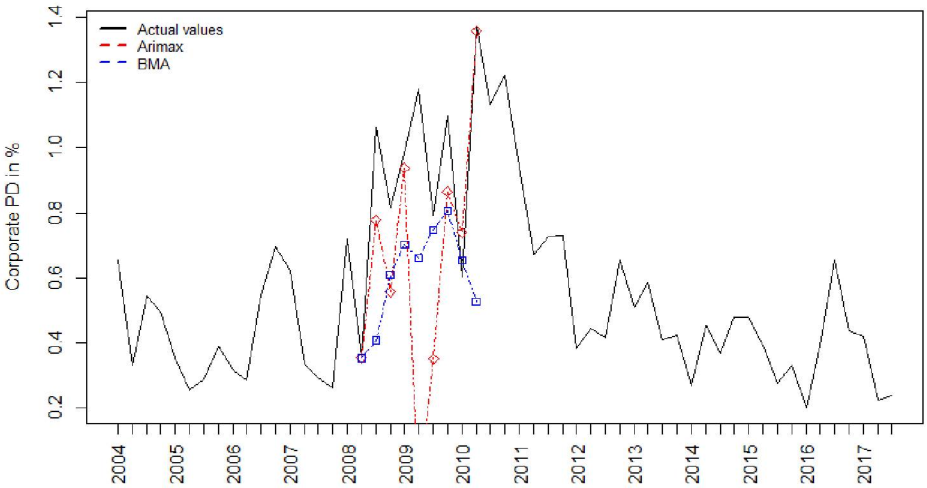


Figure 4.10:
BMA - Testing period 2008 predictions



4.1.5 Neural Networks

Last model setup to compete against Arimax setup is the Neural Network design. The crucial problem in Neural Network training process is the decision about the number of hidden layers and the number of hidden neurons. As a common practice in Economy, one hidden layer is sufficient, more hidden layers don't improve the performance and only result in overfitting. On the other hand, the number of hidden neurons in the one hidden layer is not that obvious. Too few hidden neurons results in not good enough fit. Too many neurons result in overfitting, which means that neurons don't have enough data points to train themselves. This is very crucial in short data series, such as in our case. It will be even more crucial in testing the period of financial crises after 2008 when we will have even shorter data series. All variables are standardized before the training process in order for the algorithm to converge more easily. This is done by deducting the sample mean and then dividing by the square root of the sample variance for each of the variables. Final predictions are then constructed in the same manner, with opposite direction. First candidate setup is including the same 6 variables as in the BMA setup, plus their lags up to the 4th order as well as lags of the PD. With time trend included, it results again to 35 predictors. Number of hidden layers is set to 1 and the number of hidden neurons is set to a number of predictors plus number of outputs, divided by 2 and rounded, which results into 18 hidden neurons.
Next 3 setups will use reduced number of predictors, as well as it will control the base number of neurons (computed in the same manner as in the previous case) by Optimal Brain Surgeon pruning. All these setups will consider predictors based on the best performing model from the all possible ARDL combinations of predictors and their lags, in similar manner is the choosing procedure for candidate models to BMA. One of the setups will consider the best ARDL structure from all setups, the second will consider the best structure from setups restricted to have no gaps. Last setup will use the the same predictors as in the original CNB Arimax model.

Full sample

Considering the goodness of fit on the the full sample, the first setup with no restrictions yield very good fit, with RMSE of 0.0320 (versus 0.1898 from CNB Arimax). MAE is 0.0225. Model with the best individual model from BMA setup allowing for gaps also perform better than Arimax on in sample perfor-

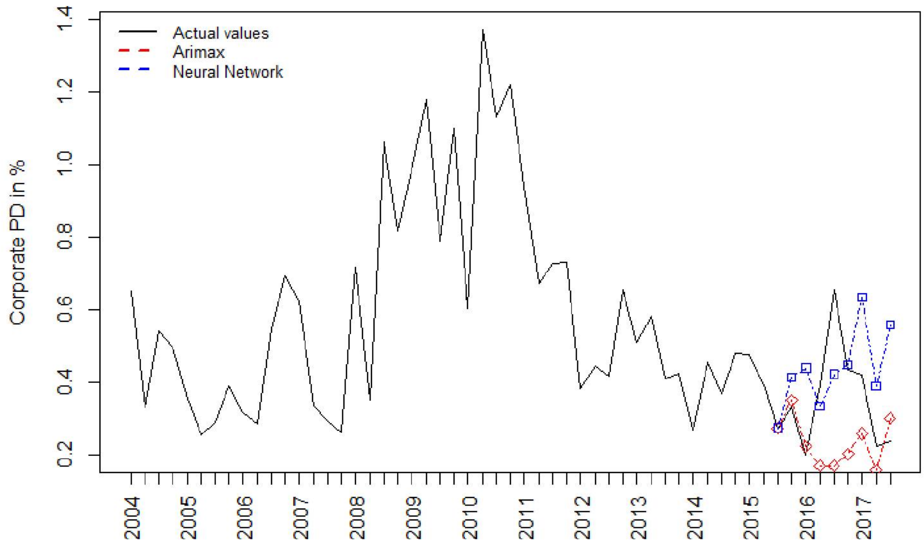
mance. It has RMSE of 0.1683 with prunning, 0.1653 without prunning. MAE is 0.1478 and 0.1318 respectively.
Considering model with no gaps restriction, it appears again that this restriction is not helping to fit better as in the case of BMA models. Prunned version has RMSE of 0.1880, non-prunned version 0.1689. MAE is 0.1372 and 0.1398 respectively.
Using the same independent variables as in the original CNB Arimax model improves the in-sample performance the most. RMSE of non-prunned Neural Networks using CNB equation is 0.0836, compared to 0.1898 of the Arimax setup. MAE is 0.0671. This particular version of variable selection can be used as a direct comparison of methods Arimax versus Neural network, since both models investigated use the same variables.

Testing period 2016

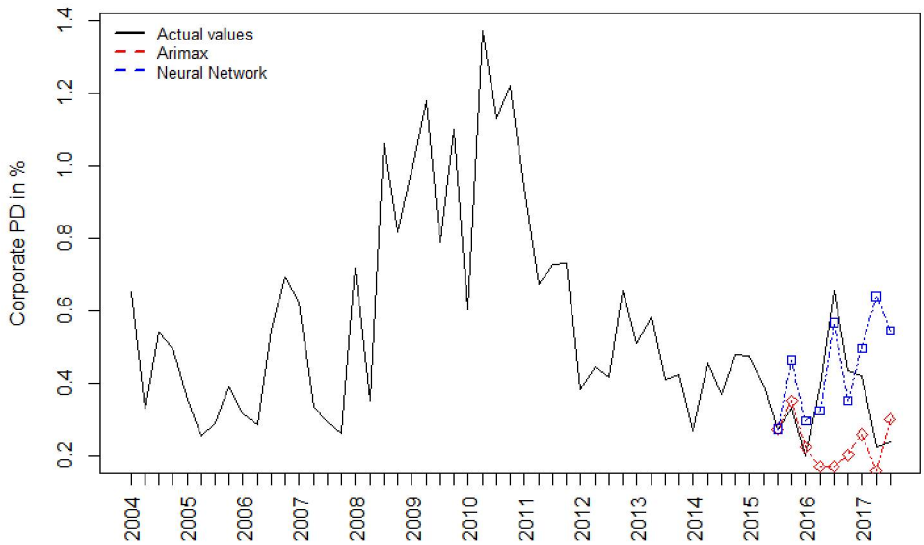
Period from Q4 of 2015 to Q3 of 2017 serves as a testing period for Neural Networks as well. Procedure is the same as in case of Arimax, FMA and BMA testing, Neural Network is trained on shorter sample up till Q3 of 2015 and predictions are then rolled for next 8 quarters. The same way as in FMA and BMA, if inputs variables are chosen based on in sample measure, they are chosen based on the shorter training sample as well.
Best performing model from full sample estimation, model with no restriction of inputs, yields very well fit in this period as well and it is the best performance among all other Neural Network options. Last known point and the next 8 predictions as well as predictions from Arimax model are in Figure 4.11 (a).

This Neural Network unlike the other methods and unlike the Arimax setup did capture the last spike in the time series, although it didn't capture it's decline. RMSE of these 8 predictions is 0.1928, MAE 0.1656.
Neural Network setup with the same variables as original Arimax model yields good results as well. Predictions are shown in Figure 4.11 (b).
In this case as well the Neural Network setup didn't predict decline of PDs, although it didn't capture the spike completely. RMSE of these 8 predictions is 0.1998, MAE is 0.1584.
Both of these Neural Network setups performed significantly better than Arimax model during this testing period.

Figure 4.11:
NN - Testing period 2016 predictions

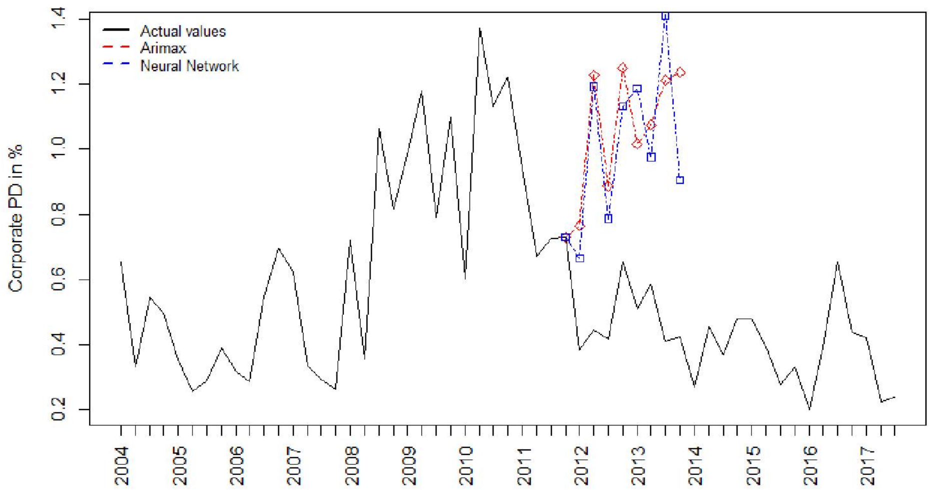


(a) Model without inputs restriction



(b) Model using CNB variables

Figure 4.12:
NN - Testing period 2012 predictions



Model using CNB variables

Testing period 2012

Unlike in Testing period 2016, in this period Neural Network setup without variable restriction didn't capture the PD decline in observed period. Predictions are very close to the Arimax setup. RMSE is almost the same as in case of Arimax, it is 0.6253, MAE is 0.6072.

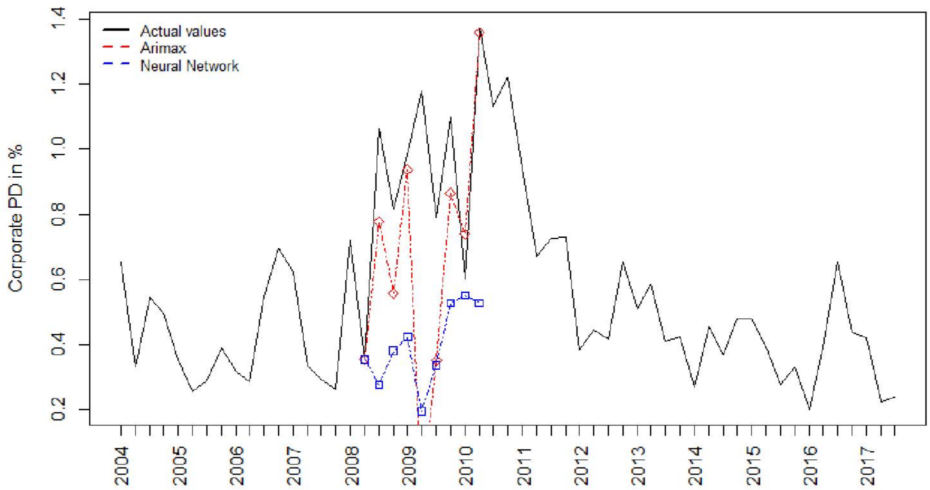
Slightly better results yield the setup with CNB variables. Even this one however didn't catch the PD decline. RMSE of these 8 quarters is 0.5960, MAE 0.5526. Predictions can be seen in Figure 4.12.

Testing period 2008

Testing period 2008 tells similar story as period 2012. Neural network with unrestricted input space didn't perform well. Performance is this time significantly worse than Arimax model. RMSE of 8 predictions is 0.5618, MAE 0.4829.

Model with CNB inputs doesn't outperform original Arimax as well. Even though predictions are stable, they don't capture high spikes starting in year 2008 at all. Predictions are in Figure 4.13. RMSE is 0.6453, MAE 0.5851.

Figure 4.13:
NN - Testing period 2008 predictions



Model using CNB variables

Neural Network estimations during testing periods 2008 as well as 2012 suffer from short data series as well. Even with one single layer, number of parameters to be estimated is huge. Even if the Neural Network had 1 hidden layer with only 2 hidden neurons, it would be necessary to estimate 18 parameters in case of CNB variables inputs. This fact suggest that Neural Network is not suitable tool for estimating models on small data sets.

4.1.6 Performance comparison for Corporate PD

This section discuss performance comparison across all estimated methods and models within methods dealing with Corporate PD. Original Arimax model didn't perform poorly in all cases as expected. It can be however outperformed by FMA, BMA and Neural Network models. This doesn't hold for the Testing period 2008, where more complicated models struggled to train themselves from very few data points, less complicated models performed similarly as Arimax model at best. Original Arimax model has however one advantage over other models considering Testing period 2008. Most of other models not only trained themselves on small data sample, but chose inputs or individual equations based on the small data sample as well. Arimax inputs are fixed during all testing

Table 4.5:
Corporate PD - RMSE comparison

RMSE	Full Sample	2016	2012	2008
CNB Arimax	0.1898	0.2160	0.6246	0.4990
FMA - No restriction	0.1548	0.1892	0.5923	
FMA - ECB selection, no gaps	0.1813	0.2835	0,6440	0.4377**
FMA - ECB seleciton, gaps allowed	0.1339*	0.1868*	0.5106*	0.5097
BMA - No restriction	0.1455*	0.1703	0.2330**	0.6079
BMA - ECB selection, no gaps	0.1857	0.1558**	0.6351	0.5811
BMA - ECB selection, gaps allowed	0.1898	0.1894	0.5049	0.4505*
NN - No restriction	0.0320**	0.1928*	0.6253	0.5618*
NN - CNB variables	0.0836	0.1998	0.5960*	0.6453

This table shows comparison of chosen models across the methods. It shows in sample RMSE on full sample and out of sample RMSE during testing periods, each consisting 8 quarters of predictions.
* indicates the best model among the same estimation framework within one period
** indicates the best model among all models within one period

periods, but variables and the method are chosen on longer data set than data until year 2008.

RMSE and MAE measures across all methods for chosen models are in Tables 4.5 and 4.6. On the full data sample, the best model according both RMSE and MAE appeared to be Neural Network model without inputs restriction, second best best model overall on full sample was Neural Network model with CNB variables. These models however don't perform that well in testing periods, which suggests that these models are over-fitted. This coresponds with the problem discussed earlier, in Neural Network design there are many parameters to be estimated, in very short data samples even more parameters than data points themselves. This suggests that Neural Network design has limited utility on small data samples.

The best model according to both RMSE and MAE from other methods than Neural Networks on the full sample is FMA model with ECB selection procedure of individual models, allowing for gaps between lags. This has significantly better fit than both the best individual candidate models, but than original Arimax model as well. Other two FMA models listed have better fit according to both RMSE and MAE than Arimax model as well as their best individual models. Improvement over original Arimax model by FMA with ECB model selection with no gaps restriction is not very significant.

All three listed models from BMA class have better fit than Arimax model

Table 4.6:
Corporate PD - MAE comparison

MAE	Full Sample	2016	2012	2008
CNB Arimax	0.1560	0.1590	0.6039	0.3336**
FMA - No restriction	0.1233	0.1594	0.5706	
FMA - ECB selection, no gaps	0.1412	0.2209	0.6394	0.3628*
FMA - ECB seleciton, gaps allowed	0.1080*	0.1551*	0,4950*	0.3704
BMA - No restriction	0.1170*	0.1401	0.2200**	0.5593
BMA - ECB selection, no gaps	0.1436	0,1187**	0.6298	0.5331
BMA - ECB selection, gaps allowed	0,1276	0.1543	0.4903	0.3616*
NN - No restriction	0.0225**	0.1656*	0.6072	0.4829*
NN - CNB variables	0.0671	0.1998	0.5526*	0.5851

This table shows comparison of chosen models across the methods. It shows in sample MAE on full sample and out of sample MAE during testing periods, each consisting 8 quarters of predictions.
* indicates the best model among the same estimation framework within one period
** indicates the best model among all models within one period

according to both RMSE and MAE. Models using ECB selection however have only marginally better fit. The best model among BMA models is BMA without candidate models selection restriction. It is comparable with the second best FMA model, without candidate models selection restriction.
Testing period 2016 shows different results, although the original Arimax model is outperformed by most of the chosen models and all of the methods. The best RMSE and MAE yields the BMA model with ECB selection. This result however stems from the fact that most of other models either didn't catch the spike during the observed period or they have some points far from the true value, even if they caught the spike. This BMA model doesn't have very dynamic predictions in all of the observed periods, it is a steady line from the last point. In this period this model is the "least wrong" one by predicting all points in the middle of the spike. Both other BMA models however produced reasonable predictions that are better than original Arimax. Two FMA models outperformed the Arimax model according to RMSE, but according to MAE they are very close to the Arimax, therefore we can't conclude with certainty that they performed better. Both Neural Networks chosen setups are outperformed by BMA and FMA models. They still yield better RMSE than Arimax during Testing period 2016, however according to MAE, they perform even worse than Arimax.
According to Testing period 2012, one model shines among other, it is BMA

model without model selection restriction. It is the only model that didn't predict another high value spikes during observed period, it predicted stagnation. Both RMSE and MAE is by far the best among chosen models. FMA and BMA models with ECB selection procedure allowing for gaps between lags outperformed Arimax model as well. They didn't predict another inflated PDs in observed period, predictions were however more stable and increase wasn't that sharp. Both Neural Network models with the remaining FMA and BMA models don't yield significantly better results than Arimax model and their performance was insufficient during observed period.
Testing period 2008 suffers from very short training period and as it could be expected, more simple model leads to better results. Simple original Arimax has the best MAE among all models during this period, it is challenged only by FMA model with ECB selection with no gaps restriction and BMA model with ECB selection allowing for gaps. All other models performed poorly.
Overall high flexibility showed the BMA model without individual model selection restriction as well as both other BMA models and FMA model with ECB selection allowing for gaps. These models in conclusion outperform original Arimax model in multiple measures. All Neural Networks models fits well on the full sample size, but they are not flexible when it comes to out of sample measurement, suggesting that Neural Network isn't suitable method in this particular case.

4.2 Housing loans

This section discusses prediction of Probability of default for Housing loans sector. Dependent variable is a level of 3M Probability of default in Housing sector measured quarterly. Data available are for period Q3 of 2008 to Q3 of 2017, which results into 41 data points. All models used in this thesis are allowing for up to 4th lag of PD, which effectively cuts out 4 data points from the time series, resulting into 37 points in total.

4.2.1 CNB Methodology Framework

Original CNB methodology for housing Probability of default is estimated by Arimax model in similar manner as corporate loans PD. Dependent variable is 3M PD in levels. Independent variables in the original model are similar as in corporate model, the lag structure is slightly different. Predictiors are: 1st and

4th lag of Probability of default, current value of 3M Pribor, current value of first difference of CZK/EUR exchange rate and its second lag and 4th lag of real GDP YoY growth. Compared to predictors of corporate model, only lags of Pribor are left out and 1st lag of PD is included. Otherwise the regresssion is the same.

Full sample

Table 4.7 shows comparison of coefficients and standard errors from simple Arimax regression on the full sample. The first column shows coefficients published in Gersl (2012). Second column shows results of re-estimated model on data up to Q3 of 2017.
Re-estimated model has more recent data and more datapoints, older model however has some older data not available, therefore intersection of the data set is less than 30 periods. Unlike in corporate model, both models have this time coefficients closer to each other. In both cases the 1st lag of PD is highly significant predictor with high coefficient, the coefficient is lower than in the original model, standard errors are similar. The 4th lag of PD has negative sign in both cases, both coefficients insignificant. Current value of 3M Pribor has in both cases coefficient close to the 5% significance threshold, in the re-estimated model the coefficient becomes significant and stronger in negative values. Current value of the first difference of CZK/EUR exchange rate far from 5% significance in both cases. Unlike in original model, second lag of the exchange rate difference is not significant. YoY GDP growth 4th lag is marginally significant in original model, but highly insignificant in the re-estimated model.

Figure 4.14 shows fitted values of the re-estimated model. In-sample RMSE is 0.0408 and MAE 0.0322. These values will serve as a baseline for comparison with other methods used on full data sample.

Testing period 2016

The same way as in corporate section, the same Arimax model is used for testing the out of sample performance during Testing period 2016, which is the same 8 quarters period as in corporate section, from Q4 of 2015 to Q3 of 2017. Model is re-estimated on shorter training period, this time consisting 29 observation. Predictions are then made for next 8 periods and compared to actual values. While constructing predictions, actual values of other variables than PD are taken, when considering PD, lags are considered as predictions

Figure 4.14:
CNB Framework - Fitted values

This graph shows fitted values of the original Arimax model (green circles) re-estimated on the full sample versus the data (solid line) for Housing segment

Table 4.7:
Housing Probability of Default

Dependent variable 3M PD_t	(1) CNB model 2012	(2) Re-estimated model
Constant	0.352*	0.740***
	(0.145)	(0.189)
3M PD_{t-1}	0.881***	0.575***
	(0.134)	(0.125)
3M PD_{t-4}	-0.184	-0.040
	(0.103)	(0.070)
3M $Pribor_t$	-0.032	-0.081*
	(0.018)	(0.035)
$\Delta CZK/EUR_t$	0.023	-0.010
	(0.020)	(0.010)
$\Delta CZK/EUR_{t-2}$	0.046*	0.006
	(0.020)	(0.009)
GDP YoY $growth_{t-4}$	-0.014*	0.071
	(0.007)	(0.315)
N	30	37
Adjusted R2	0.911	0.953

This table shows comparison of coefficients of Household PD model between original model from Gersl (2012) and re-estimated model for purpose of this thesis. Standard Errors in parenthesis. Significance signs: *5%, **1%, ***0.1%

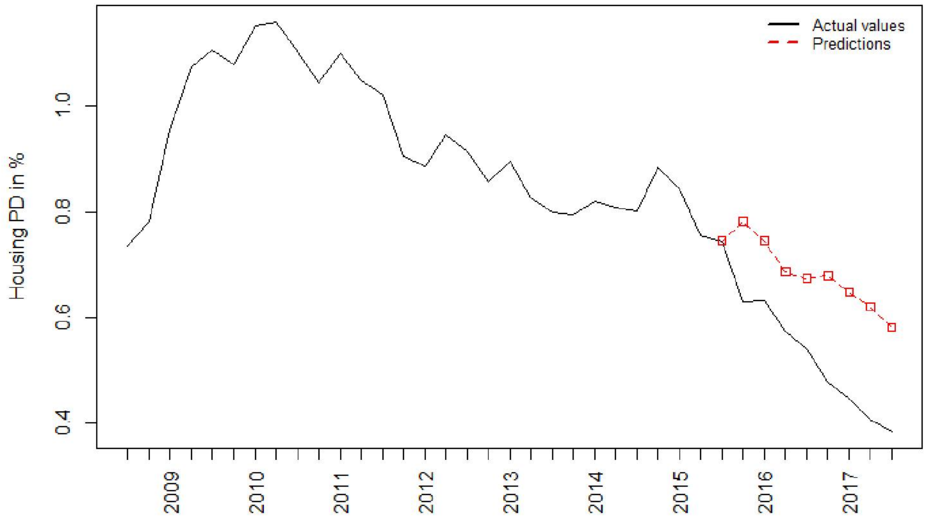
themselves. Last known point and 8 following predictions are visualised in Figure 4.15.

The first prediction expected increase of PD instead of its decline, from that point the decline was in line with the actual decline, but with weaker trend. RMSE of these 8 predictions is 0.1702 and MAE is 0.1656. These results will serve as a baseline for comparison with other methods for this training period.

Testing period 2014

Since time series for Housing PD is noticeably shorter than in case of corporate sector, it is impossible to construct the same testing periods as in corporate case, in particular the Testing period 2008, since the first quarter of that Testing period is Q3 of 2008, which is the first point of the whole data sample for Housing sector. To maintain 3 testing periods in addition to the same Testing periods 2016 and 2012, one period of 2014 is introduced to replace testing period 2008. Testing period consists 8 quarters from Q4 of 2013 to Q3 of 2015. Original Arimax model is re-estimated on training set pre Q4 2013, which

Figure 4.15:
CNB Framework - Testing period 2016 predictions



counts 21 periods. Last known point and 8 predictions from Arimax model are in Figure 4.16.

Arimax model during this testing period predicted steady decline of PD with not much other movements and missed the Actual values including one more peak at Q4 of 2017. We can see that here is definitely space for some improvement in performance. RMSE of these 8 predictions is 0.1852, MAE 0.1657.

Testing period 2012

Last Testing period is period from Q1 of 2012 till Q4 of 2013 and it is the same Testing period as in corporate section labeled as Testing period 2012. This period suffers even more from small data sample than the period 2008 from corporate section. We have available only 14 data points prior this testing period. Last known point and 8 quarters are in Figure 4.17.

Even though the model doesn't predict PD decline immediately in the first predicted quarter, it expected huge decline in next periods, getting close to zero in the last period. RMSE on 8 quarters of predictions is 0.2207, MAE 0.1923.

Figure 4.16:
CNB Framework - Testing period 2014 predictions

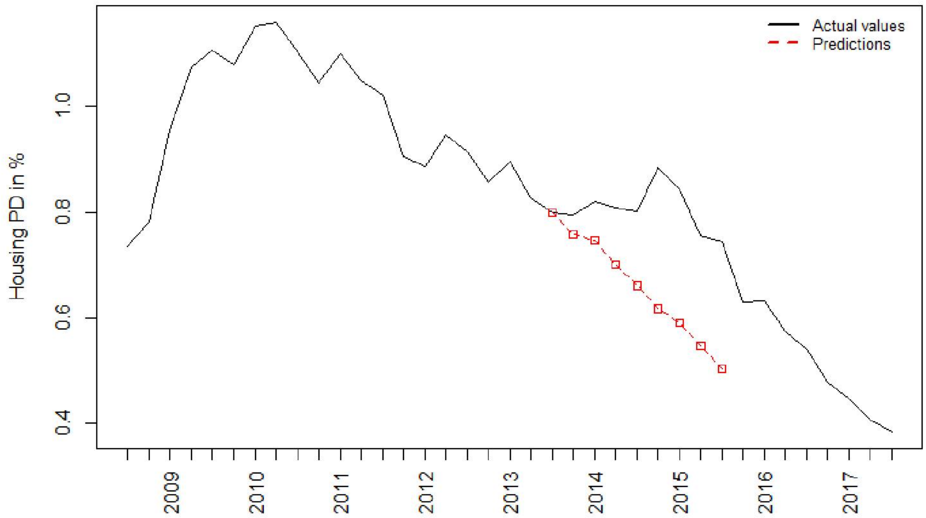
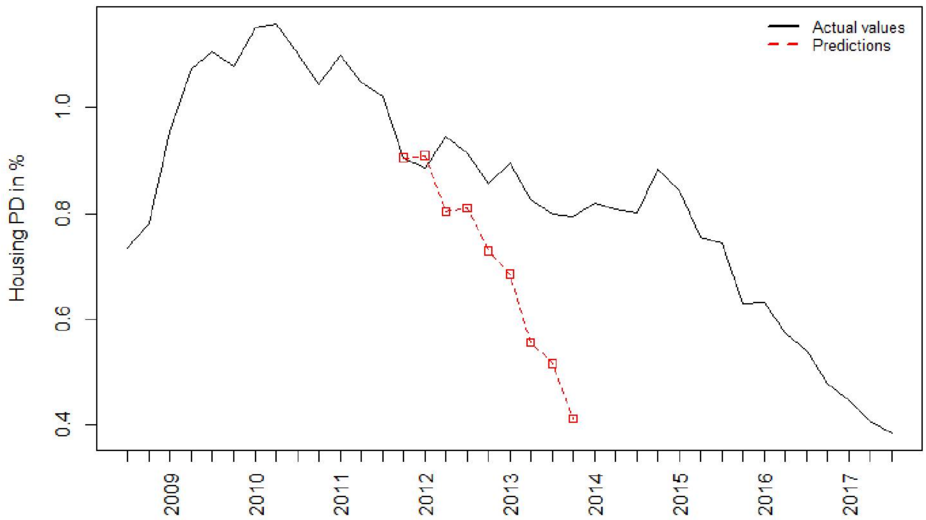


Figure 4.17:
CNB Framework - Testing period 2012 predictions



4.2.2 Forecast combination

The same approach used in corporate section, motivated by Papadopulos (2016), shall be used in Housing sector as well. In corporate sector the hypothesis by Papadopoulos isn't confirmed, hypothesis of better performance of combined model over best non-combined models. In addition to independent variables used in corporate section, one more is used, being Property Price in levels, all equations include time variable to deal with a trend. Strategy is the same as before, to estimate linear model with all possible combinations of subsets of 11 predictors, forming $2^{11} = 2048$ possible models.

Next step is to keep only models that have all variables significant on 5% level. This forms 53 surviving models. The most frequent variable is Inflation, present in 27 models. Least frequent variable is Compensation of employees present in 6 models and CZK/EUR exchange rate in 9 models. All other variables are in 10 to 16 models.

For all 53 models we obtain predictions. To combine these predictions we use again 3 weighting schemes, average predictions, median predictions and Bates-Granger weights. Besides the option to keep all 53 models, we can trim out some poorly performing models, which is done by sorting them by RMSE and cutting out some portion of the worts models. Weights are then assigned to the new surviving model space.

At the end of the procedure, we have 53 (or less in case of trimming) individual predictions and 3 combined prediction vectors per depth of trimming.

Now we can investigate performance of each version. The best model according to RMSE is model with independent variables being Pribor, GDP, Debt to GDP ratio, Household consumption and Unemployment. This individual linear model has RMSE of 0.0416. The worst model from the full space of 53 models has RMSE of 0.1197, average is 0.6236. Bates-Granger weighted model has RMSE of 0.0451, mean weights 0.0493 and median predictions 0.0473. Once again, this doesn't support the hypothesis of Papadopoulos, where all weighting schemes performed better than the best individual model. Even the best Bates-Granger scheme has worse RMSE than 3 best individual models.

The same conclusion is in case of MAE. The best model has MAE 0.0321, average MAE in full model space is 0.0504. MAE under Bates-Granger weights (calculated by MAE) is 0.0369, under mean weighting 0.0406, median weighting 0.0389. This time is the Bates-Granger scheme worse than 6 individual models.

If we trim some of the candidate models and keep only the best performing once, we can obtain better results. Keeping only 10 best models improves the combined model, but it still doens't perform better than the best individual model according to RMSE. Bates-Granger weights scheme yields RMSE 0.0419, mean weighting 0.0420, median 0.0429. This time however all the weighting schemes perform better than all of the individual models except the best one. Considering MAE on the trimmed model space, the mean MAE is 0.0363. MAE under Bates-Granger sheme (weights according to MAE) is 0.0320, which is already better than the best individual model. Mean weighting produce better MAE as well, being 0.0321. Median weights doesn't beat the best model with MAE of 0.0330.
These results suggest that the full space of candidate surviving models can't outperform the best individual model after combining them together. Trimmed space combination however performs slightly better than the best individual model according to MAE, worse according to RMSE, therefore the performance is ambiguous.

4.2.3 Frequentist Model Averaging

In corporate section, first model averaging technique to challenge original Arimax model demonstrated its power, because model combinations not only outperformed original Arimax models, but best individual models in among candidate equations as well. For Housing PD similar strategy is used. First FMA model follows the Forecast Averaging procedure by estimating all possible subset regressions from given set of variables and their lags. Second step is keeping only regressions that have all coefficients significant on 5% level. Surviving models then have their coefficients weighted and final predictions can be constructed.
Second option is to leave out the 5% significance restriction on all variables. This improved the performance in the corporate sector models. Both of these options encounter problem with the size of the equation space. Once again, the set of predictors is reduced, keeping: Pribor up to a 2nd lag, Unemployment up to a 2nd lag, GDP growth up to a 4th lag, difference of exchange rate up to a second lag and lagged values of PD up to a 4th lag. This forms 18 possible predictors and therefore $2^{18} = 262144$ equations.
Other option is again to use the ECB selection procedure. From a set of non-lagged predictors, each subset is examined and the optimal lag structure is

found by minimizing the Bayesian Information Criterion. 2^n quations, where n is the number of non-lagged predictors, then form the candidate model space from where the weighted coefficients are taken. For this type of model (and for similar models in BMA section) these 6 variables are chosen: Pribor, Inflation, Unemployment, YoY GDP growth and QoQ Property Price growth. Time variable is added to each equation as well. Both version with no gaps restriction and without this restriction shall be examined.

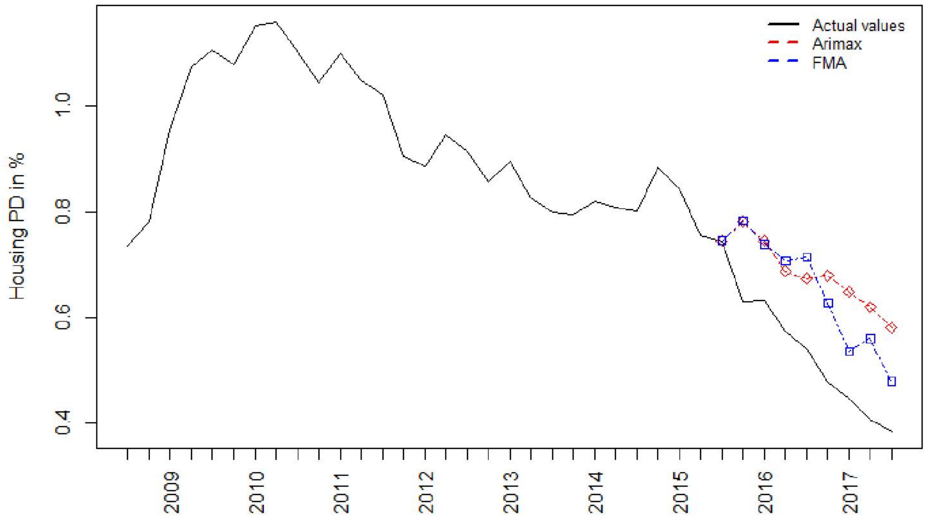
Full sample

The first setup examining 262144 equations results into 84 equations with all variables significant on 5% level. The best equation according to AIC and BIC is equation with the current value of Unemployment and its 2nd lag, 1st and 4th lag of GDP growth. Unemployment and its 2nd lag are as well the most frequent variables, both being in 37 models. The least frequent variable are lags of exchange rate. In sample RMSE of the combined model is 0.0356 (versus 0.0408 of the original Arimax), the best individual model has almost the same RMSE of 0.0356. MAE of the combined model is 0.0278, best individual model 0.0277. Unlike in the corporate sector, these results suggest that the combined model is at best performing similarly as the best individual model.

In corporate sector relaxing the significance condition improved performance of the model. This fact is supported in Housing sector as well. Keeping all 262144 and constructing predictions from them improves the in sample RMSE, which is now 0.0327, however the best individual model in this setup has RMSE lower, 0.0314. MAE is 0.0251 and 0.0242 respectively. Therefore even in this case, combined model doesn't outperform the best individual model, even though it fits better than original Arimax model.

Next option of introducing the ECB selection procedure with no gaps condition doesn't support the combination idea as well. Combined without gaps has RMSE of 0.0316, the best individual model 0.0310, MAE is 0.0255 versus 0.0256. According to these measures the performance comparison is ambiguous. If we however relax the no gaps condition we obtain once again better results. Combined model in this case has RMSE of 0.0284 versus 0.0290. MAE is 0.0224 versus 0.0229. This model fits better than the best individual model, better than original Arimax and better than all other estimated FMA models.

Figure 4.18:
FMA - Testing period 2016 predictions



Testing period 2016

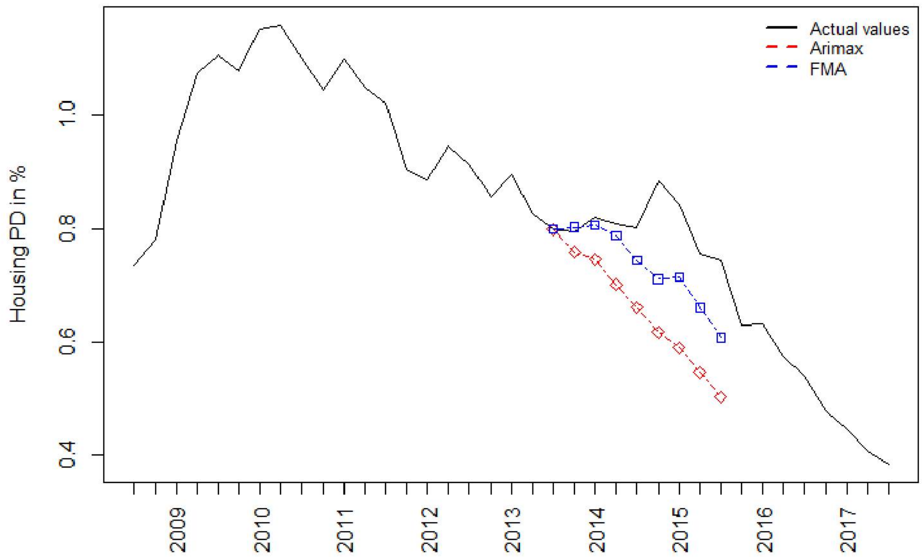
Period between Q4 of 2015 to Q3 of 2017 consisting of 8 quarters serves as the first tool for out of sample performance. Both models without ECB selection procedure performs poorly, similarly as the original Arimax at best. Models with ECB selection outperform Arimax model slightly, both with very similar predictions. Last known point and 8 predictions of the model without no gaps restriction are in Figure 4.18.

During the first periods of the testing window both Arimax and FMA predicted similar values, with decline of PD above the Actual values. In the second half of the testing period, decline of predictions by FMA accelerated. RMSE of FMA is 0.1349 versus 0.1702 of Arimax. MAE is 0.1318 versus 0.1656. Results of the model with no gaps restrictions are similar. Both models slightly outperform Arimax model.

Testing period 2014

Similar takeaways are results of the Testing period 2014. Both FMA without ECB individual model selections missed with predictions. FMA with ECB selection and with no gaps condition produced almost the same predictions as Arimax model. But as in the previous testing period, model with ECB

Figure 4.19:
FMA - Testing period 2014 predictions



selection without no gaps condition performs better than Arimax. Predictions are in Figure 4.19.

As opposed to Arimax model, this FMA setup didn't predict decline of PD in first periods, however it expects similar decline as Arimax model in later periods. RMSE of these 8 periods is 0.0986 versus 0.1852 of Arimax. MAE is 0.0788 versus 0.1657.

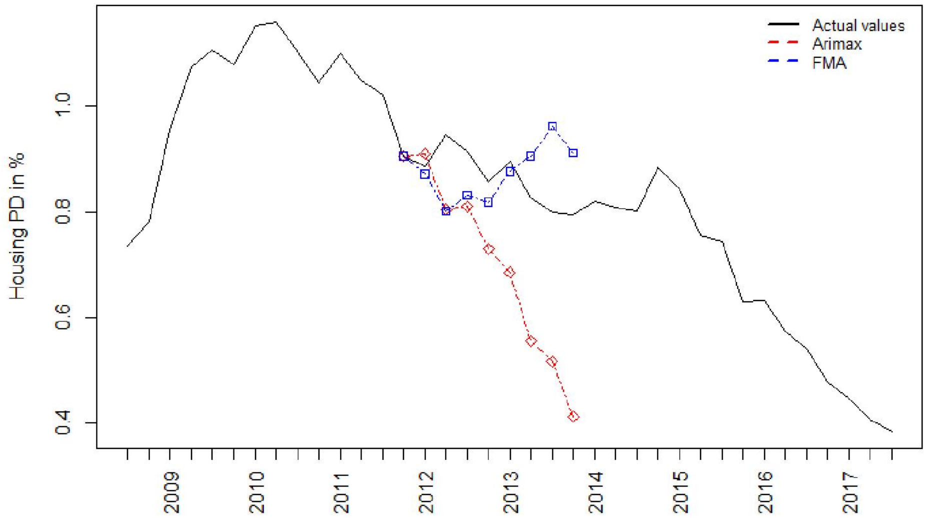
Testing period 2012

Last testing period suggests once again similar results. Both FMA models without ECB procedure this time perform even worse than original Arimax model. On the other hand both FMA models predict better. Once again, the FMA with ECB procedure without no gaps restricion predicts the closest results to Actual values, predictions compared to Arimax model are in Figure 4.20.

In the beginning the FMA model expected similar evolution as Arimax model, however PDs then recovered to higher values. RMSE is 0.0975 versus 0.2207 of Arima, MAE 0.0822 versus 0.1923.

As a conclusion, models without restrictions on candidate model selection sim-

Figure 4.20:
FMA - Testing period 2012 predictions



ilar to ECB procedure don't perform better than Arimax model across all testing periods. As opposed to that, models with ECB procedure predict better, in particular the model without no gaps restriction, which is the best one regardless on the testing period chosen.

4.2.4 Bayesian Model Averaging

Main competing model against Arimax is again the current framework of ECB. Dependent variable is again transformed by the logit transformation, which keeps the predictions bounded to interval (0, 1). Prediction are then transformed back using sigmoid function.

This section uses the same 6 independent variables as in the FMA model: 3M Pribor, Property Price QoQ growth, Inflation, Unemployment, YoY GDP growth, first difference of CZK/ EUR exchange rate. These 6 predictors and their lagged values up to the 4th order together with lags of PD to the 4th order and time variable form 35 potential predictors to the BMA procedure.

Full sample

The same way as in FMA, for all possible combinations of 6 predictors the optimal lag structure is found by minimizing the Bayesian Information Crite-

rion. This forms $2^6 = 64$ candidate equations to be formed. ECB methodology requires to consider only equations without gaps. As we've seen in previous parts of the thesis, this restriction sometimes turns out to be very restrictive and worsens the performance significantly. The same holds for the logit transformation. Not performing the logit transformation and then back the sigmoid transformation can improve the performance significantly. For these reasons we again investigate versions with the no gaps restriction, version with relaxing this condition and version with relaxing the whole candidate model selection procedure completely. All these versions shall be performed both with and without the logit transformation. This will offer us 6 model versions.
ECB selection procedure with no gaps restriction and with logit transformation turns out to be once again very restrictive. Out of all 30 possible lagged values, only 6 of them are used, mainly 1st lag of PD in 100% of candidate equations. 1st lag of GDP growth is present in 25% of candidate equations, 1st lag of inflation in 23%. Other lags, being 1st and 2nd lag of Unemployment, 1st lag of exchange rate difference and 1st lag of Pribor are in less than 20% of equations. Looking at posterior probabilities of individual predictors, 1st lag of PD has 100% inclusion probability, 1st lag of GDP growth has 82.5% inclusion probability, Pribor 33%, the rest of lagged values have less than 6% inclusion probability. The best equations are mostly based on several non-lagged variables. Best 5 individual models (based on model posterior probability) have between 5 and 8 predictors, which is more than in corporate sector. These best 5 models account for 74.3% cumulative posterior probability. In sample RMSE of this model is 0.0338, which is better than original Arimax 0.0408. MAE is 0.0278
If the no gaps restriction is relaxed, way more lagged values make it into the model, it is 24 lagged variables in at least one candidate ARDL structure. Interestingly lagged variable with the highest posterior inclusion probability from the no gaps model, 1st lag of GDP growth, has lost most of it's predictive power, in this setup it has only 0.5% posterior probability. 1st lag of PD has no longer posterior probability of 100%, it drops down to 42.6%. Highest posterior probability now lies on the 4th lag of Unemployment with probability 82%. The best models again rely on larger number of predictors than the same model estimated for corporate sector, number of variables is between 6 and 8 for the best 5 models. Cumulative posterior probability of the best 5 models is 81%. Unlike in corporate sector, relaxing the no gaps condition didn't improve the in sample performance this time, RMSE is 0.0393, MAE 0.0301.

Taking into account model not using the ECB candidate model selection and considering all possible combinations of models includes all variables and their lags, each with probability at least 1%. Highest inclusion posterior probability is on current values of Inflation and Unemployment, with probability 100%, followed by 2nd lag of Pribor and 3rd lag of Inflation and 3rd lag of Unemployment with posterior probabilities 98%, 97% and 92%. This setup improves in sample performance significantly, RMSE is 0.0244, MAE 0.0186.

In corporate sector estimation without the logit transformation improved the performance. It turns out that in the housing sector it is not always the case. ECB selection with no gaps condition uses more lagged variables than its logit version, 9 instead of 6, although all the lagged variables except the 1st lag of PD are present in less than 8% of candidate equations. The best equations are this time individual models with less variables, between 3 and 8 for the best 5 models. Cumulative posterior probability is lower for these 5 best models, being 49%. Performance is worse than in case of its logit counterpart, RMSE is 0.0364, MAE 0.0285.

Relaxing both logit transformation and no gaps condition gives better results than its logit counterpart. This model uses as well as its logit version a lot of lagged variables, 25 in at least one candidate model. The highest posterior probability lies on the 3rd and 4th lag of Unemployment, current value of Unemployment and Inflation, all these being above 90%. The best 5 models use between 3 and 7 predictiors, cumulative probability of these 5 models is 97%. RMSE of this version is 0.0378, MAE 0.0293, which suggests that it is performing better than its logit counterpart, but worse than logit version of ECB selection with no gaps condition.

Model without logit transformation and without using ECB selection procedure is performs better than versions with ECB selection procedure, however slightly worse than the same model without logit transformation. Again, all the variables are used, this time most of them with more than 2% inclusion probability. Highest inclusion probability yields current value of Unemployment, its 3rd lag and 3rd lag of Property price QoQ growth, all these 3 with higher inclusion probability than 98%. RMSE is 0.0248, MAE 0.0203, both measures worse than its logit counterpart.

Corporate BMA models on full sample suggested that logit transformation doesn't improve performance as well as more restrictions on model selection don't. This doesn't hold for Housing sector, where the effect of logit transformation and selection restrictions is ambiguous.

Figure 4.21:
BMA - Testing period 2016 predictions

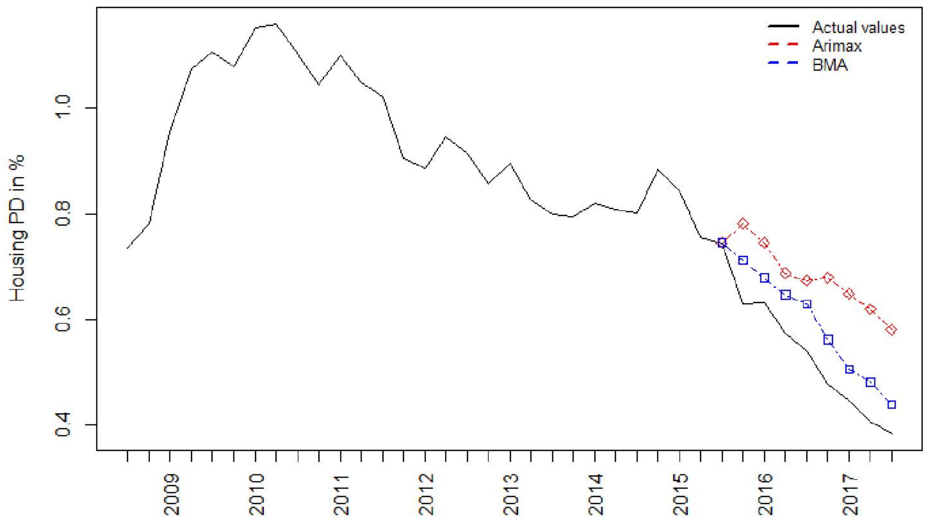


Table 4.8 shows inclusion probabilities of individual variables of three models. Column (1) logit transformation with ECB selection with no gaps restriction, Column (2) no logit transformation, ECB selection with relaxed no gaps condition, (3) logit transformation with no candidate models selection restrictions.

These results go against the variable selection into original Arimax model. Except the 1st lag of PD, all other variables included in the original Arimax (4th lag of PD, current value of Pribor, current value of exchange rate and its 2nd lag, 4th lag of GDP growth) doesn't have big support in BMA models. Crucial role seems to play Unemployment and its lags, which supports the hypothesis that Unemployment is crucial predictor. Inflation is included significantly as well.

Testing period 2016

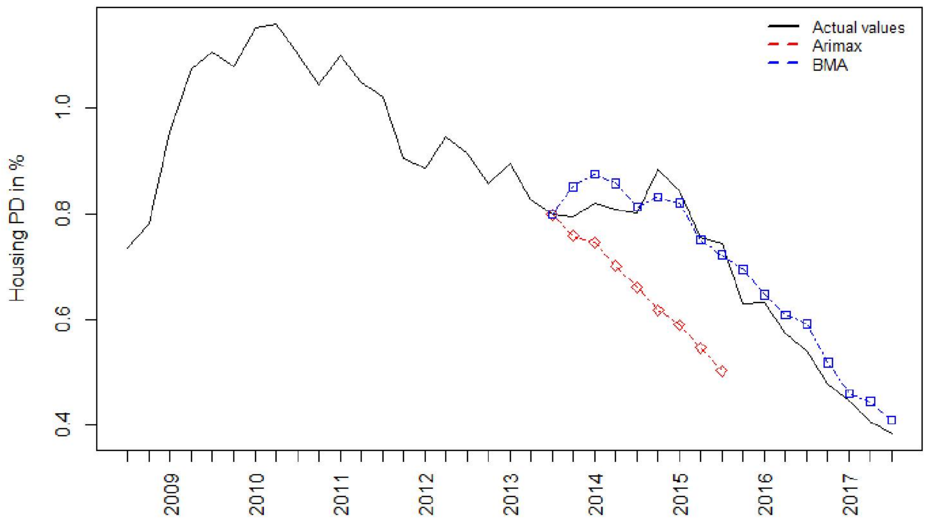
The best model from the full sample comparison performs as the best in Testing period 2016 consisting quarters Q4 of 2015 to Q3 of 2017 as well. As the only model, BMA without restrictions but with the logit transformation did predict very close to the true decline, with similar dynamics. Comparison of last known point and 8 quarters of predictions compared to original Arimax predictions is in Figure 4.21.

Table 4.8:
BMA posterior inclusion probabilities
Housing sector

	(1)	(2)	(3)
Intercept	100.0	100.0	100.0
3M Pribor$_t$	38.2	11.3	1.3
Inflation$_t$	93.7	97.8	100.0
Unemployment$_t$	90.8	95.3	100.0
GDP YoY Growth$_t$	84.1	0.1	8.4
PP growth QoQ$_t$	19.9	80.2	2.5
$\Delta CZK/EUR_t$	21.8	2.8	1.8
PD$_{t-1}$	100.0	18.3	34.1
PD$_{t-2}$			75.0
PD$_{t-3}$			2.7
PD$_{t-4}$			16.1
3M Pribor$_{t-1}$	33.2		7.2
3M Pribor$_{t-2}$		7.4	97.5
3M Pribor$_{t-3}$			1.8
3M Pribor$_{t-4}$			1.9
Inflation$_{t-1}$	5.4		23.3
Inflation$_{t-2}$			3.5
Inflation$_{t-3}$		12.7	97.2
Inflation$_{t-4}$		81.7	3.3
Unemployment$_{t-1}$	0.8	0.1	3.4
Unemployment$_{t-2}$	0.8	0.1	9.7
Unemploymentt-3		95.0	92.0
Unemploymentt-4		92.4	86.8
GDP YoY Growtht-1	82.5	0.7	27.5
GDP YoY Growtht-2			14.0
GDP YoY Growtht-3			7.8
GDP YoY Growtht-4			8.3
PP growth QoQt-1		0.1	7.3
PP growth QoQt-2			1.5
PP growth QoQt-3		82.2	19.6
PP growth QoQt-4			20.5
$\Delta CZK/EURt-1$	0.1		1.6
$\Delta CZK/EURt-2$			2.4
$\Delta CZK/EURt-3$		2.0	1.9
$\Delta CZK/EURt-4$		0.2	1.8

This table shows Posterior Inclusion Probabilities of individual predictors for BMA.
Column (1) logit transformation, ECB selection, no gaps
Column (2) no logit transformation, ECB selection, gaps allowed
Column (3) logit transformation, no selection restrictions

Figure 4.22:
BMA - Testing period 2014 predictions



RMSE of these 8 predictions is 0.0720 (versus 0.1702 of Arimax), MAE is 0.0706. Other BMA models however didn't catch the decline.

Testing period 2014

Testing period 2014 from Q4 of 2013 to Q3 of 2015 suggests similar results. This time again BMA without model selection restriction performed very well. Predictions compared to Arimax predictions are in Figure 4.22.

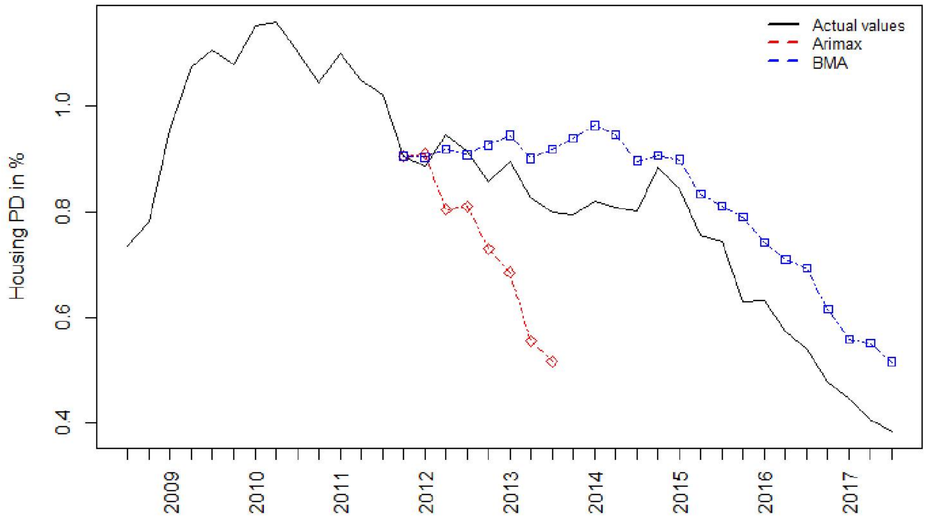
This model captured both stagnation in the beginning of the testing period and then even decline at the right time. RMSE of 8 predictions is 0.0397 (versus 0.1852 of Arimax), MAE is 0.03426. All other BMA setups performed better than original Arimax, but they don't copy the stagnation and decline as close as above mentioned BMA.

Testing period 2012

The same model as according to the full sample and both testing periods 2016 and 2014 performs remarkably even in this testing period from Q1 2012 to Q4 of 2013. Predictions compared to Arimax model are in Figure 4.23.

The model not only correctly predicted the stagnation in the beginning of the observed period, but then even decline approximately at the right time.

Figure 4.23:
BMA - Testing period 2012 predictions



RMSE of these 8 predictions is 0.0786 (versus 0.2207 of Arimax), MAE is 0.0634. Most of other BMA models however performed poorly. Most of them predicted sharp decline and then next spikes at the levels of the highest peak in 2010.

Overall, BMA models mostly performed better than Arimax model in all testing periods, some of them however performed poorly anyway. The best model by far among BMA models is the same model with logit transformation without model selection restriction across all testing periods.

4.2.5 Neural Networks

Last model framework to compete against original Arimax are Neural Networks. In the corporate section Neural Networks didn't perform very well, especially in short training samples. Since in Housing sector we have available even less data points, similar conclusions can be expected.

Four types of models shall be investigated. First one being model with the same 6 predictors as in BMA and most of FMA setups and their lags up to a 4th order. These 35 inputs won't be restricted any further.

Second option is inspired by the ECB selection procedure. The best model according from the candidate models with no gaps restriction is chosen by

minimizing BIC. Its independent variables serve as inputs to the Neural Network. Both inputs and hidden neurons are either controlled by the Optimal Brain Surgeon prunning or not prunned. Next option is to use the same selection, but relax the no gaps restriction, model is estimated again both with and without prunning.
Last option is to use the same variables as the original Arimax uses as inputs to Neural Network. This provides a direct comparison of these two methods.

Full sample

Similarly to the corporate sector, model with unrestricted inputs, which counts 35 in total, fits very well, however it is expected that the model is overfitted and won't perform very well on testing samples. In sample RMSE is 0.0138, versus 0.0408 of Arimax.
As in corporate sector, model restriction to models with inputs without gaps doesn't improve fitting. Prunned version has RMSE 0.0482, MAE 0.0370, which is worse fit than Arimax model. Non-prunned version improves RMSE to 0.0400, MAE 0.0311.
Relaxing the non gaps restriction further improves the performance, RMSE is 0.0365 and MAE 0.0292 in prunned model and 0.0327 and 0.0245 respectively for non-prunned version.
Using inputs from the original Arimax model fits the model better than Arimax itself, RMSE is 0.0380, MAE 0.0305.

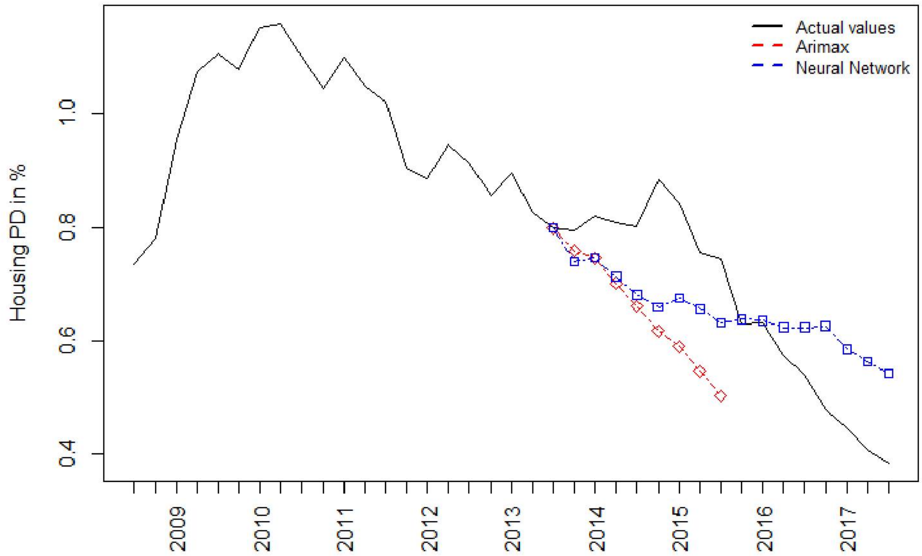
Testing period 2016

None of the Neural Networks setting predicts better than original Arimax model does during Testing period 2016. Some of them predict even less steep decline of PD. The best settings, being model without input restrictions and both prunned and non-prunned version of ECB selection without no gaps restricting, achieve similar results as Arimax model. RMSE of ECB selection with allowed gaps without prunning has RMSE of 0.1714 (vs 0.1702 of Arimax) and MAE 0.1674.

Testing period 2014

All of the Neural Networks predicted similar evolution as Figure 4.24, which shows predictions of not restricted Neural Network.

Figure 4.24:
NN - Testing period 2014 predictions



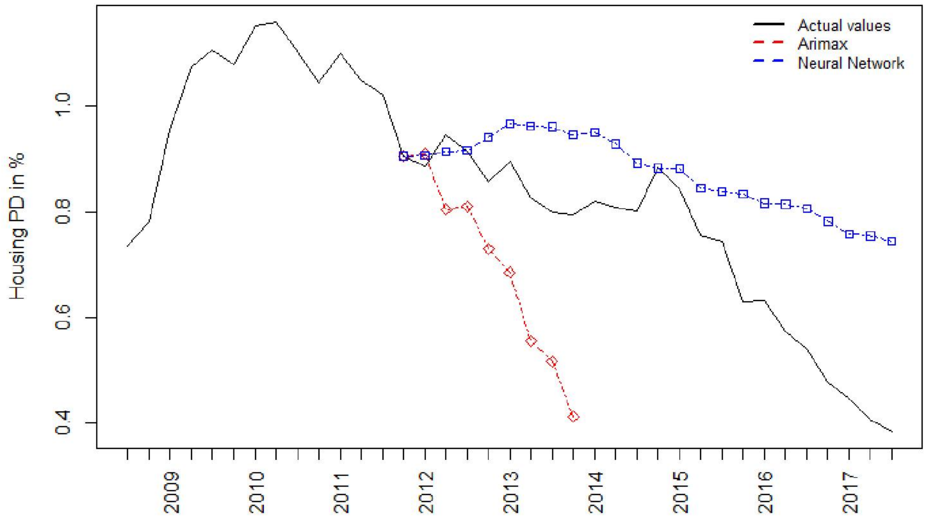
Prediction like this can't be really called a success, even though the prediction is closer to the Actual value than Arimax model. RMSE of this model is 0.1288 (versus 0.1852 of Arimax), MAE is 0.1183.

Testing period 2012

Best performance among Neural Networks after learning from the shortest data sample is non-prunned version of ECB selection without no gaps restrictions. Predictions are in Figure 4.24. Even though the model didn't expect sharp decline as Arimax model, in long run the predictions are not flexible. RMSE of 8 predictions is 0.1006, MAE 0.0823. Other Neural Network models predict similar evolutions.

Overall most of the Neural Network models performed similarly, better than Arimax during testing periods 2012 and 2014, worse or the same at best during period 2016. Due to the short time series are predictions however not flexible and we can expect that in case of a stress period, Neural Networks would underestimate the shock.

Figure 4.25:
NN - Testing period 2012 predictions



4.2.6 Performance comparison for Housing PD

This section discuss performance comparison of models dealing with Housing Probability of default across all discussed methods for chosen models. Unlike in the corporate sector, original Arimax model performed poorly during all testing periods, showing very little flexibility and not dynamic or accurate predictions. Most of the discussed models from all estimation frameworks were able to outperform Arimax model in in sample measures on full data sample and according to out of sample measures during three testing periods, this time being periods around years 2016, 2014 and 2012. Comparison of RMSE and MAE across all methods and all testing periods for chosen models is in Tables 4.9 and 4.10.

The best model performing model from this sectors is without doubts the BMA model with logit transformation without individual model selection motivated by ECB. This model has both the best MAE and RMSE across all periods and in full sample as well. It also graphically predicts the true values as well as the dynamics with precision. Both other chosen BMA models perform significantly better than Arimax as well, in particular in testing periods 2012 and 2014.

Table 4.9:
Housing PD - RMSE comparison

RMSE	Full Sample	2016	2014	2012
CNB Arimax	0.0408	0.1702	0.1852	0.2207
FMA - ECB selection, no gaps	0.0316	0.1185*	0.1781	0.1743
FMA - ECB seleciton, gaps allowed	0.0284*	0.1349	0.0986*	0.0975*
BMA - No restriction, logit	0.0245**	0.0720**	0.0796**	0.0786**
BMA - ECB selection, no gaps, logit	0.0338	0.1652	0.1272	0.1656
BMA - ECB selection, gaps allowed	0.0380	0.1381	0.1623	0.0933
NN - ECB selection, gaps allowed	0.0327*	0.1714*	0.1403	0.1006*
NN - CNB variables	0.0380	0.2448	0.1397*	0.1055

This table shows comparison of chosen models across the methods. It shows in sample RMSE on full sample and out of sample RMSE during testing periods, each consisting 8 quarters of predictions.
* indicates the best model among the same estimation framework within one period
** indicates the best model among all models within one period

Table 4.10:
Housing PD - MAE comparison

MAE	Full Sample	2016	2014	2012
CNB Arimax	0.0322	0.1656	0.1657	0.1923
FMA - ECB selection, no gaps	0.0255	0.1137*	0.1555	0.1488
FMA - ECB seleciton, gaps allowed	0.0224*	0.1318	0.0788*	0.0822*
BMA - No restriction, logit	0.0185**	0.0706**	0.0343**	0.0633**
BMA - ECB selection, no gaps, logit	0.0277	0.1652	0.1093	0.1411
BMA - ECB selection, gaps allowed	0.0293	0.1350	0.1381	0.0796
NN - ECB selection, gaps allowed	0.0245*	0.1674*	0.1283	0.0823*
NN - CNB variables	0.0305	0.2314	0.1169*	0.0890

This table shows comparison of chosen models across the methods. It shows in sample MAE on full sample and out of sample MAE during testing periods, each consisting 8 quarters of predictions.
* indicates the best model among the same estimation framework within one period
** indicates the best model among all models within one period

The second best model from the full sample space is FMA model with ECB selection relaxing the no gaps restriction. It also performs very well in periods 2014 and 2012, it however doesn't catch the right trajectory as good as BMA model. Second chosen FMA model with ECB selection and no gaps restriction performs well according to the full sample and testing period 2016, but in periods 2014 and 2012 it predicts poorly.

Neural Network models don't suffer from short time series as much as in corporate case, moreover when the available PD data for Housing sector provide significantly less data points. Most of the Neural Network schemes performed similarly to each other and better than Arimax model according to RMSE and MAE, exception is the newest Testing period 2016, where all the models performed either similarly or worse than the Arimax model. Neural Networks in this sector showed however very little flexibility and it is very likely that they would underestimate the risk in case of an adverse scenario.

4.3 Consumer loans

Consumer loans are the last discussed credit risk sector in this thesis. Data for the Consumer loans sector have the same length as for the Housing sector, from Q3 of 2007 to Q3 of 2017, which is 41 quarters of data. As in the previous two sectors, models are estimated for up to a 4th lag of PD, which effectively cuts out 4 more data points, therefore we have 37 data points available.

4.3.1 CNB Methodology Framework

Original CNB methodology for the Consumer sector is modeling the first difference of Consumer Probability of default measured quarterly. This comes from the findings of the original paper about non-stationary behaviour of the time series. It can be pointed out that previous two sectors have non-stationary behaviour of PD as well, however to keep in line and to compare results with CNB Arimax models, it is necessary to keep the structure of variables the same as in the original paper. Taking the first difference effectively cuts out one more training data point, therefore it only 36 data points on PD are available.

Original Consumer Arimax model uses fewer predictiors than other two sectors. It uses only 4 independent predictors from which two of them are 3rd and 4th lag of PD first difference. Only other two predictors are 2nd lag of real GDP Quarter on Quarter growth and 4th lag of Quarter on Quarter Property price

Table 4.11:
Consumer Probability of Default

Dependent variable $\Delta 3MPD_t$	(1) CNB model 2012	(2) Re-estimated model
Constant	-0.009	0.051
	(0.020)	(0.036)
$\Delta 3MPD_{t-3}$	0.356*	-0.190
	(0.152)	(0.161)
$\Delta 3MPD_{t-4}$	0.055	-0.130
	(0.157)	(0.160)
QoQ GDP growth$_{t-2}$	-4.489***	-0.330
	(1.744)	(0.281)
QoQ Property Price growth$_{t-4}$	0.018***	0.0804
	(0.004)	(1.381)
N	30	36
Adjusted R2	0.652	0.014

This table shows comparison of coefficients of Consumer loan PD model between original model from Gersl (2012) and re-estimated model for purpose of this thesis. Standard Errors in parenthesis. Significance signs: *5%, **1%, ***0.1%

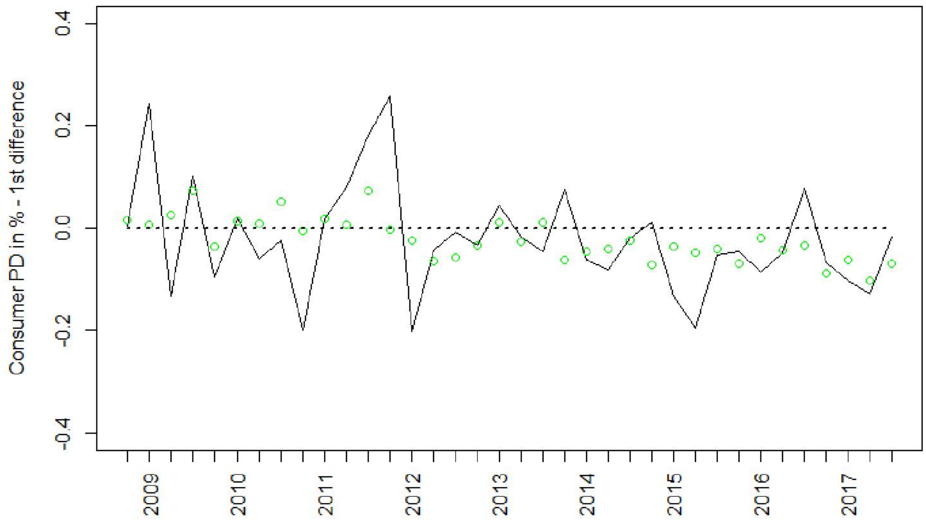
growth.
Table 4.11 compares coefficients of the original model published in Gersl (2012) and re-estimated model on new data.

Unlike in corporate or housing sector, this time we were not successful in reproducing the equation to have similarities. Coefficients are different both in sizes and significances. Adjusted R squared is much lower as well. As in case of housing sector even this time the re-estimated model has more recent data and more data points, however the older model has some older data not available now, therefore the intersection of both data sets is less than 30 data points. Fitted values are in figure 4.26. RMSE of the whole fit is 0.0971, MAE 0.0699. These measures will serve as a comparison of full sample fit with other methods.

Testing period 2016

As in corporate and housing sector, first testing period is period of 8 quarters from Q4 of 2015 to Q3 of 2017. Training period prior Q4 of 2015 contains 28 of observations. The same way as in previous two sectors, 8 quarters of predictions are constructed in a way that variables except PD are taken as true values, lags of PD are considered prediction themselves. Last known point and 8 quarters of predictions are visualised on Figure 4.27.

Figure 4.26:
CNB Framework - Fitted values



This graph shows fitted values of the original Arimax model (green circles) re-estimated on the full sample versus the data (solid line) for Consumer segment

Figure 4.27:
CNB Framework - Testing period 2016 predictions

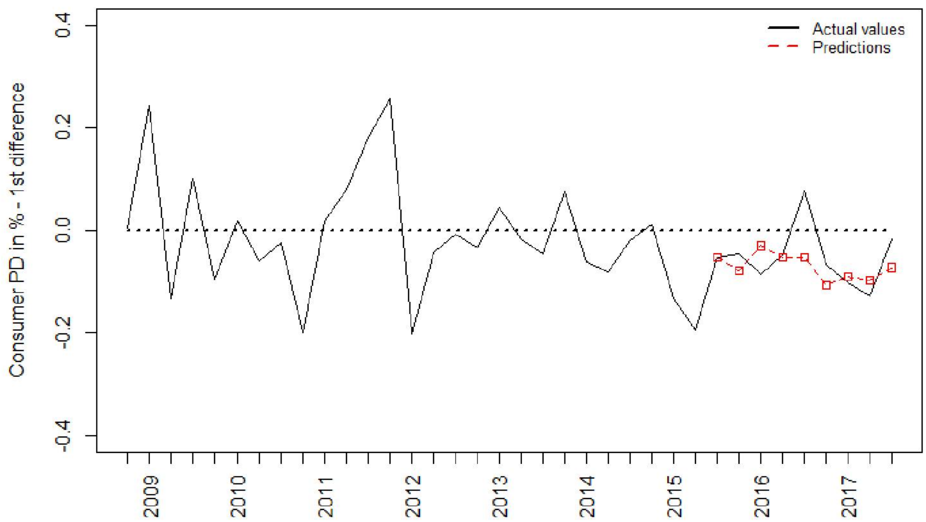
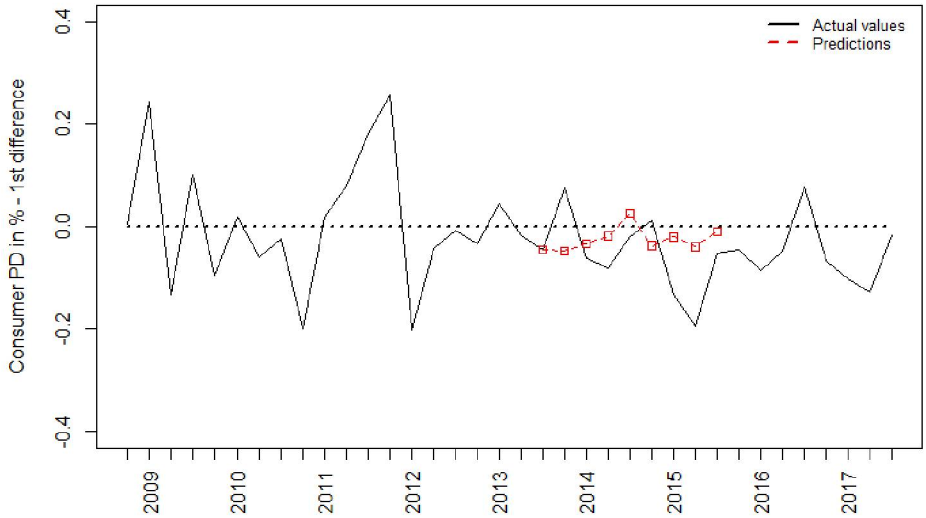


Figure 4.28:
CNB Framework - Testing period 2014 predictions



We can see that predictions don't copy Actual values well at all . In contrary, predictions don't get close to any actual movement. RMSE of 8 predictions is 0.0585, MAE 0.0454. This shall serve as a baseline for comparison of other models during Testing period 2016.

Testing period 2014

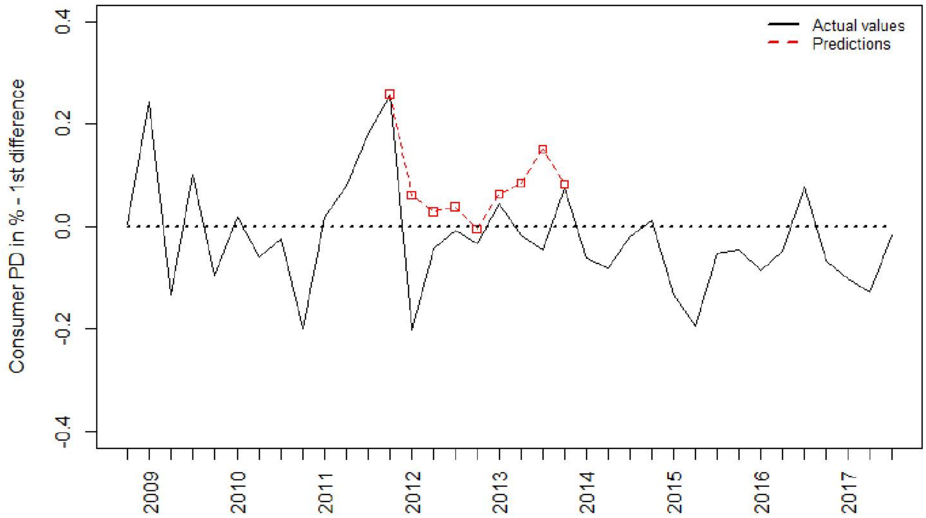
The same way as in the housing sector, it is not possible to test performances on periods 2008, 2012 and 2016 due to short time series. Period between Q4 of 2013 and Q3 of 2015 replace the Testing period 2008. Performance of the model is poor, with similar conclusions as in the previous period. Predictions are shown in Figure 4.28. RMSE of 8 predictions is 0.0882, MAE 0.0770.

4.3.2 Testing period 2012

Testing period 2012 is expected to suffer from short training sample prior the testing period even more than in case of housing sector. Training sample prior Q1 of 2012 contains only 13 data points. Last known point and 8 quarters of predictions are in Figure 4.29.

Among all the testing period, these predictions can be considered as the

Figure 4.29:
CNB Framework - Testing period 2012 predictions



closest ones to the actual values. They still however don't perform sufficiently well, predictions are often with the opposite sign than the Actual values, predictions increase of the PD in levels instead of decline. RMSE of 8 predictions is 0.1256, MAE is 0.0921.

4.3.3 Forecast combination

Approach from Papadopoulos (2016) shall be replicated as well in Consumer sector to challenge the hypothesis of the authors paper that even in model without lags, while combining predictions from more candidate models, predicts better than the best individual model, across various weighting schemes. The procedure estimates all possible linear models from all combinations of predictors. Only models with all variables significant on 5% are then considered to the model combination. Consumer sector here have an issue with significancy. In Corporate and Housing sector, there were 32 and 53 models respectively tagged as surviving models with all variables significant. This procedure in case of Consumer sector has only 2 fully significant individual models, combining only 2 models cease to have a logic. Instead that all the models regardless on significancy are taken into account. Predictors for the procedure are Pribor, Real GDP in levels, CZK/EUR exchange rate in levels, Debt to GDP

ratio, Household Consumption, Compensation of employees, Inflation, Unemployment, QoQ GDP growth, QoQ Property Price growth and Consumption to GDP. This forms 2048 individual equations, not further restricted.
Even here in the Consumer loan sector, results don't support the authors hypothesis. The best individual model have RMSE of 0.0821. Combined RMSE with Bates-Granger weights on the full space of models have RMSE 0.0917, mean predictions 0.0923, median prediction 0.0934. Similar results are in case of MAE, the best individual model has 0.0624, combined model with Bates-Granger weights. Exactly same conclusions are while considering MAE.
In previous sectors, the more the model space was trimmed, the better results were obtained. In Housing sector the combined model had better MAE than the best sector while trimming the individual model space. In Consumer sector however, even trimming to best 5 five models doesn't help the combined model to achieve better fit than the best individual model. Bates-Granger weights obtain 0.0823 RMSE, mean weighting almost identical, median weights slightly higher 0.0827. Similar results are takeways from MAE comparison.
Even Consumer sector doesn't support the authors hypothesis that combined model even with no lags can fit better than best individual model. It also doesn't support the hypothesis that less complicated weighting schemes yield better results than more complicated ones.

4.3.4 Frequentist Model Averaging

First model framework to challenge the original CNB Arimax model is FMA. In previous two sectors, slightly different results from each other were obtained. In the corporate sector various FMA models outperformed not only original Arimax models, but they outperformed the best individual models as well. The same doesn't always hold for the housing sector, although some models yield decent results beating both Arimax and the best individual models.
The same strategy for the third time shall be used in the Consumer sector. Two different strategies are applied once again. First one takes vast range of variables and lags and similarly to the Forecast combination section, all possible linear models are estimated, these model then serve as individual candidate equations, where their coefficients are weighted to obtain final combined coefficients.
Second option is to borrow the ECB model selection procedure by finding best optimal ARDL structure for each subset of potential non-lagged variables,

which forms 2^n equations, where n is the number of non-lagged predictors. Both schemes with no gaps restriction and with this restriction relaxed shall be examined. Performance improvement connected to the no gaps restriction is ambiguous in previous two sectors.

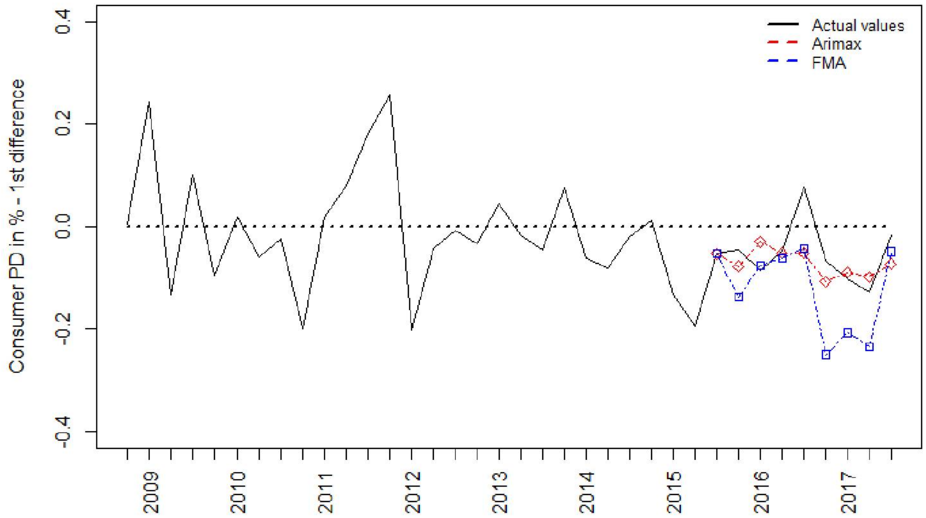
Full sample

First option with not restricted individual model selection has once again problem with the model space size. For this section these variables were chosen as potential predictors: Unemployment and its lags up to the 2nd order, QoQ GDP growth up to a 2nd lag, QoQ Property Price growth up to a 4th lag, first difference of CZK/EUR exchange rate and lags of PD up to a 4th order. This forms as in previous sectors $2^{18} = 262144$ individual linear equations to be estimated. For the same reasons as in the Forecast combination section, in this sector the model space can't be reduced by keeping only linear models with all variables significant, because of the issues with low significancy of variables in this sector. Therefore only model combining all 262144 equations is considered. In sample RMSE of this model is 0.0745, whereas the best individual model has RMSE of 0.0811 and original Arimax model 0.0971. MAE suggests similar results, combined model has 0.0572, whereas the best individual model 0.0644 and Arimax 0.0699. This FMA therefore fits better than both original Arimax model and the best individual model.

Second option is to use the ECB selection procedure. Independent variables chosen for this model are Pribor, Inflation, Unemployment, QoQ GDP growth, QoQ Property Price growth and first difference of CZK/EUR exchange rate. This forms $2^6 = 64$ candidate equations, each with optimal structure. Combined model with no gaps restriction has RMSE of 0.0761, which is better than original Arimax, however worse than the best individual model with RMSE of 0.0758. MAE of the combined model is 0.0571 versus 0.0616 of the best individual model. Therefore it is not certain, which model performs better, even though both fit better than original Arimax.

Relaxing the no gaps restriction improves the performance significantly. Combined model has RMSE of 0.0596 versus 0.0631 of the best individual model. MAE points out the same fact, combined model has MAE 0.0451 and the best individual model has 0.0494.

Figure 4.30:
FMA - Testing period 2012 predictions



Testing period 2016

Even though all three discussed FMA models fit better on the full sample than the original Arimax model, they most certainly don't perform well in the Testing period 2016. All of them have even worse performance than the original Arimax model. The best model is the unrestricted model, predictions are in Figure 4.30.

RMSE of these 8 predictions is 0.1002, MAE 0.0830, which is in both cases worse than the original Arimax model.

Testing period 2014

Similar results are in case of the Testing period 2014. All three models failed to provide reasonable results. The best one is model with ECB selection without no gaps restriction. RMSE of 8 predictions is 0.1138, MAE 0.0930, which is still significantly worse than original Arimax model.

Testing period 2012

Testing period 2012 isn't exception among testing period of FMA in Consumer sector. All FMA models performed poorly, the best being again model with

ECB selection procedure with RMSE, but even this one stays far below the original Arimax in performance.

4.3.5 Bayesian Model Averaging

Main competing model against Arimax is once again BMA. Dependent variable in this sector is not transformed by the logit transformation, since dependent variable are differences, which often have negative values. BMA models use the same 6 independent variables as FMA: Pribor, Inflation, Unemployment, QoQ GDP growth, QoQ Property Price growth, first difference of CZK/EUR exchange rate and their lags up to a 4th order, together with lags of PD. This again with time period variable forms 35 potential predictors.

Full sample

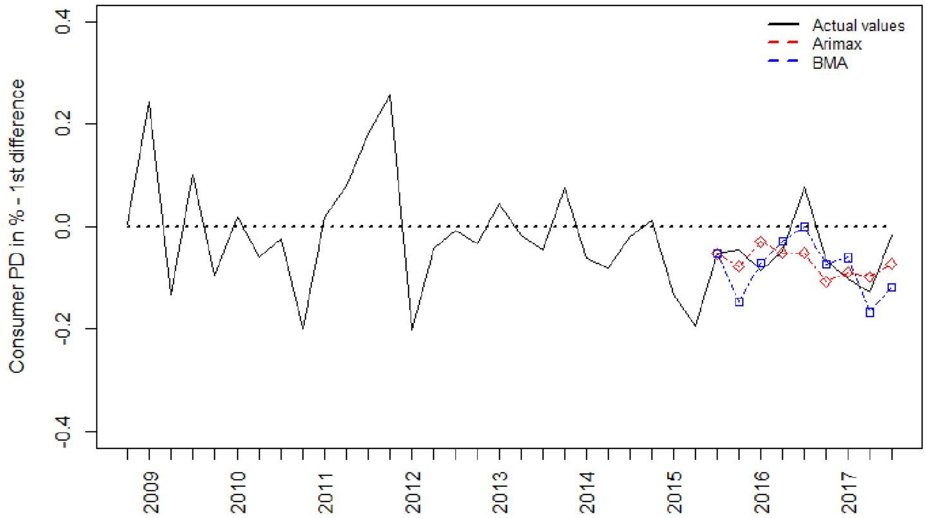
ECB selection procedure is used again. From 6 potential predictors 64 candidate equations are found by minimizing BIC for each subset of potential predictors. Once again we estimate models with the no gaps restriction as well as model relaxing this restriction. Third model shall be model without the ECB selection procedure.

In previous sectors, model with no gaps restriction turned out to be very restrictive in aspect of number of lagged variables used. In consumer sector however, model uses 12 lagged variables, which is more than in any other sector with no gaps restriction. The model however fails to fit well. The best model according to the model posterior probability of inclusion is model with only Constant. Fit is not flexible and fitted values are close to a constant. RMSE is 0.0997 (vs 0.0971 of the original Arimax), MAE 0.0734.

Model with the no gaps restriction uses even more lagged variables, 26 in particular. Highest individual posterior inclusion probabilities are on 4th lag of Inflation, 4th lag of Pribor and current value of the exchange rate difference, all of these three with posterior probability over 90%. Best five models use several variables, from 6 to 8, cumulative posterior probability of these is 87%. This model fits well as well, RMSE is 0.0640, MAE 0.0460.

Relaxing the ECB selection procedure further improves the in sample fit. Variables with the highest individual posterior probability are 3rd lag of Unemployment, 4th lag of Pribor and the current value of exchange rate difference. All of these variables have over 97% posterior probability. Since the BMA here uses more models, posterior probability is spread in more models. Best 5 models

Figure 4.31:
BMA - Testing period 2016 predictions



account for 13.7% cumulative posterior probability. This model has the best fit among BMA models with RMSE of 0.0459 and MAE 0.0386.

Testing period 2016

Both models with ECB selection strategy perform poorly in Testing period 2016. Reasonable predictions however produced model without ECB selection procedure. Predictions are visualised in Figure 4.31.

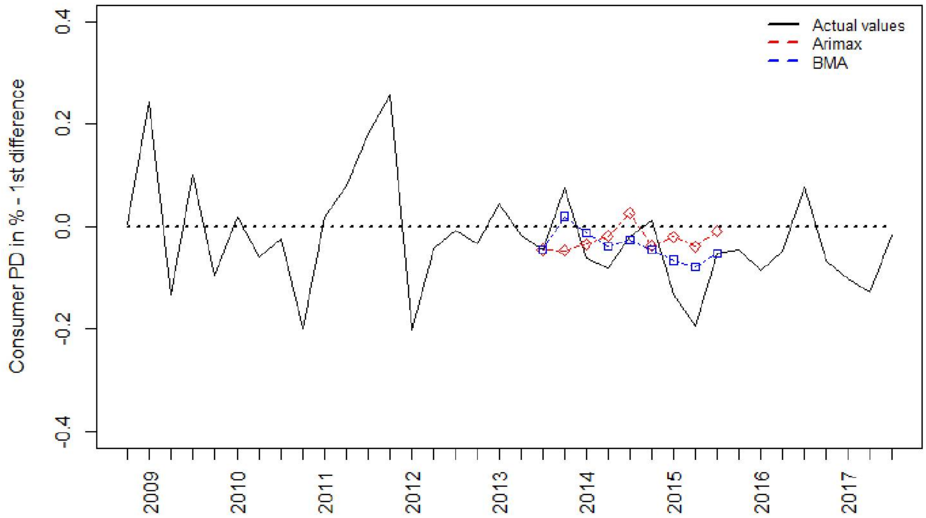
Even though RMSE and MAE of 8 predictions is worse than of the original Arimax model, graphically it performs decently. Except the first underestimated prediction, all other predictions are close to Actual movements. RMSE of 8 predictions is 0.0625, MAE 0.0507.

Testing period 2014

Testing period 2014 is similar to period 2016 in a way that both BMA models using ECB selection procedures perform poorly, significantly worse than the original Arimax. Decent performance is however present in case of BMA without ECB selection. Predictions are in Figure 4.32.

BMA model predicts well first half of the testing period, then it doesn't catch spikes up and down. RMSE is however 0.0595 (versus 0.0882 of the

Figure 4.32:
BMA - Testing period 2014 predictions



original Arimax), MAE 0.0491. BMA model here therefore outperforms the Arimax model.

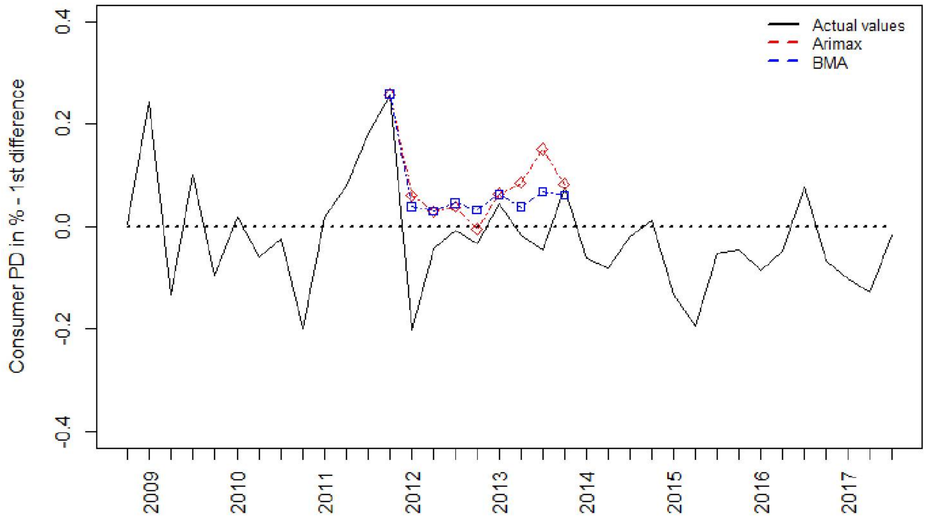
Testing period 2012

The last testing period suggest similar results than the previous ones. BMA models with ECB selection procedure perform very poorly. BMA without this procedure predicts decently.

Looking at predictions in Figure 4.33 we can see that in the beginning both Arimax and BMA predict very close to each other. Arimax model is here more dynamic, and BMA has better RMSE and MAE only because of one wrong prediction at the end of the period. RMSE of BMA is 0.1040 vs 0.1256, MAE 0.0793.

Overall BMA models performed better than FMA models. Results were consistent across all testing periods, BMA model not using ECB selection procedure performs the best among all BMA models in the Consumer segment and it performs better than Arimax in 2 periods, performance is however not that significantly better.

Figure 4.33:
BMA - Testing period 2012 predictions



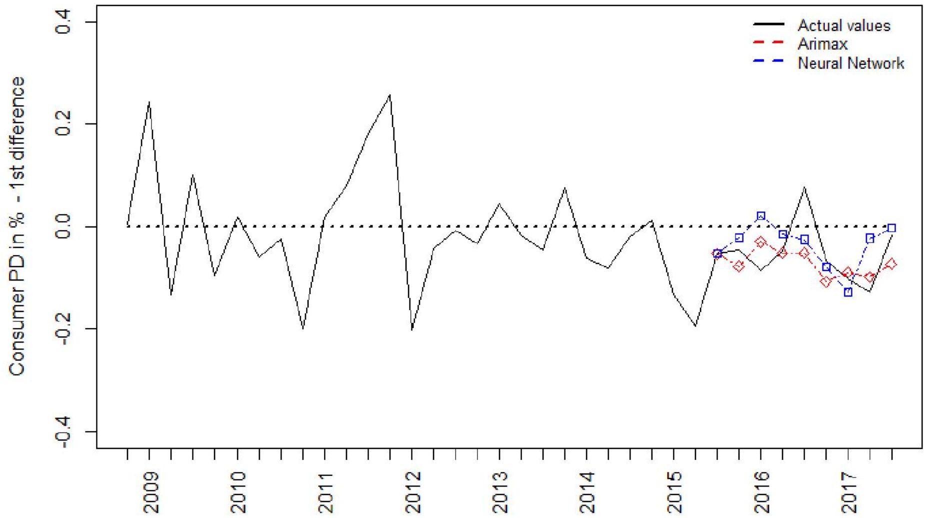
4.3.6 Neural Networks

Neural Network framework is the last model to compete against original Arimax. In Corporate segment Neural Networks didnt't perform very well. In Household sector Neural Networks predicted better, except the Testing period 2016. The same strategy will be used in Consumer sector as well. The same 6 predictors (and their lags up to a 4th order) as in BMA and most of the FMA setups shall be used. First option is not to restrict the inputs in any way and throw in all 6 variables and all their lags.
Second option is to use the best model from the ECB selection procedure (inputs to the best ARDL model according to BIC), again with and without the no gaps restriction and again in prunned and not prunned version by the Optimal Brain Surgeon.
Last option is to use Neural Network with the same inputs as are the independent variables in the original Arimax model. This can serve as a direct comparison of both methods.

Full sample

As in case of the Corporate sector, not restricted input space produce Neural Network with almost perfect fit. RMSE is 0.0006 and MAE 0.0005. This how-

Figure 4.34:
NN - Testing period 2016 predictions



ever suggests overfitted Neural Network.

In previous segments, restricting the model space to inputs without gaps didn't improve the performance. This Neural Network fits only slightly better than the original Arimax model. Prunned version has RMSE of 0.0885, MAE 0.0697, non-prunned 0.0821 and 0.0665.

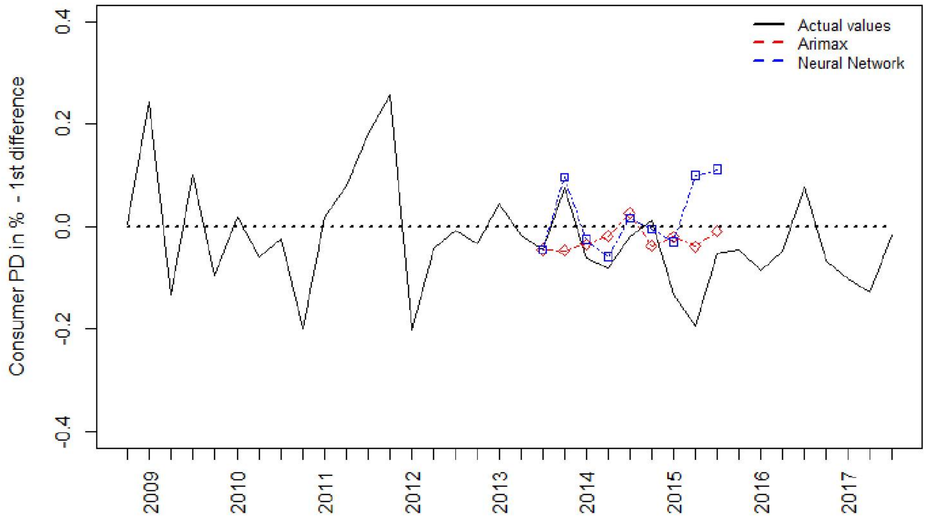
Relaxing the no gaps restriction further improves the in sample fitting. RMSE of this Neural Network is 0.0350 of the prunned version and 0.0126 of the non-prunned. MAE is 0.0291 and 0.0093 respectively.

Fit better than the original Arimax has Neural Network with the same inputs as the Arimax model. RMSE is 0.0684, MAE 0.0501.

Testing period 2016

None of the Neural Networks predict better than the original Arimax model in the Testing period 2016. Neural Networks however show some flexible and dynamic predictions unlike Arimax model. The best Neural Network is the model without model selection restriction. Predictions for this testing period are in Figure 4.34.

Figure 4.35:
NN - Testing period 2014 predictions



Although the predictions are dynamic, they don't predict well the Actual values. RMSE is 0.0674, MAE 0.0532.

Testing period 2014

In Testing period 2014, most of the Neural Networks didn't predict even close to the actual values. Neural Network with not restricted inputs however performs very well in initial periods of the testing period. Predictions however diverge far from the Actual values for the last 3 quarters of the testing period, lifting RMSE and MAE up above the levels of the original Arimax model. Predictions are in Figure 4.35. RMSE of these 8 predictions is 0.1264, MAE 0.0866.

Testing period 2012

The shortest training period consisting of only 13 data points was too short to train any Neural Network design to produce reasonable results. Most of the predictions come from the last known point in the training period, which is a global maximum of the whole series, and diverge very far from the Actual values.

Table 4.12:
Consumer PD - RMSE comparison

RMSE	Full Sample	2016	2014	2012
CNB Arimax	0.0971	0.0585**	0.0882	0.1256
FMA - No restriction	0.0745	0.1002*		
FMA - ECB seleciton, gaps allowed	0.0596*	0.1046	0.1138*	0.2902*
BMA - No restriction	0.0459*	0.0625*	0.0595**	0.1040**
BMA - ECB selection, no gaps	0.0997	0.1217	0.1164	
BMA - ECB selection, gaps allowed	0.0604	0.1118	0.1257	0.1608
NN - No restriction	0.0006**	0.0674*	0.1264*	
NN - ECB selection, gaps allowed	0.0126	0.1022	0.2749	
NN - CNB variables	0.0684	0.0716	0.2770	

This table shows comparison of chosen models across the methods. It shows in sample RMSE on full sample and out of sample RMSE during testing periods, each consisting 8 quarters of predictions.
* indicates the best model among the same estimation framework within one period
** indicates the best model among all models within one period

4.3.7 Performance comparison for Consumer PD

This section compares results across all methods and across all testing samples for chosen models. Unlike in previous two loan segments, the original Arimax model in Consumer segment hasn't been successfully replicated, model is very different from the one published in the current CNB methodology paper. Both fitted values in the whole sample as well as the predictions in testing sets from the Arimax models were not flexible, not dynamic, not able to predict more significant movements. Other model framework are often predicting dynamic predictions, but in some points far from the Actual values, which makes the predictions of the Arimax model, which are close to zero, better according to RMSE and MAE. Comparison of chosen models across all the methods are in Tables 4.12 and 4.13.

FMA models were able to fit the PDs well in the full sample, but they are all predicting poorly in across all the testing samples. In BMA, the model without ECB selection procedure is able to produce decent predictions, which are better than predictions of the original Arimax model. Exception is the Testing period 2016. Other BMA models however predicted poorly, similarly to FMA models. Neural Networks suffer from all the sectors here in the Consumer sector the most from the short training periods. Models are often overfitted. Neural Network with no restriction on the selection procedure is able to produce meaningful predictions in testing periods 2016 and 2014.

Table 4.13:
Consumer PD - MAE comparison

MAE	Full Sample	2016	2014	2012
CNB Arimax	0.0699	0.0455**	0.0770	0.0921
FMA - No restriction	0.0572	0.0830		
FMA - ECB seleciton, gaps allowed	0.0451*	0.0830*	0.0930*	0.2810*
BMA - No restriction	0.0386*	0.0507*	0.0491**	0.0793**
BMA - ECB selection, no gaps	0.0734	0.1034	0.0981	
BMA - ECB selection, gaps allowed	0.0460	0.0907	0.0770	0.1478
NN - No restriction	0.0005**	0.0532*	0.0866*	
NN - ECB selection, gaps allowed	0.0093	0.0823	0.2579	
NN - CNB variables	0.0501	0.0528	0.2656	

This table shows comparison of chosen models across the methods. It shows in sample MAE on full sample and out of sample MAE during testing periods, each consisting 8 quarters of predictions.
* indicates the best model among the same estimation framework within one period
** indicates the best model among all models within one period

Chapter 5

Conclusion

This thesis is dealing with the satellite forecasting model of Probability of Default in Czech Republic. The model serves for bank stress testing in Czech economy. In this particular context, the model is especially useful to test the resilience of the banking sector, to foresee possible recession or to construct adverse scenarios.
Credit risk satellite model currently used in Czech Economy is old and it is very likely that it wouldn't correctly predict future recessions. Current methodology doesn't use modern techniques and is fairly unchanged since year 2012.
Aim of this thesis was to propose alternative models using modern methods and compare them with the original methodology of CNB.
Dependent variable for models is 3 months Probability of Default of a subject in given segment. The original models from CNB are estimated with simple Arimax models, candidate methods to compete with their predictions are Frequentist Model Averaging, Bayesian Model Averaging (currently used in satellite models for stress testing in European Central Bank) and Neural Networks. Comparison of methods is made in 3 different segments of loans - Corporate, Housing and Consumer. Original model is re-estimated on new data and compared to competing models from proposed methods on full sample and on three testing periods, each consisting of 8 quarters.
The original Arimax model is not performing as poorly as expected in the Corporate segment, where it produces reasonable predictions. These predictions however often miss the actual values by a lot. In Housing segment the performance of original Arimax model turns out to be worse. Predictions are not dynamic and in testing periods mostly only follow the most recent trend in almost linear way. In Consumer segment re-estimated model has a lot of dis-

similarities with the originally published model from CNB. Re-estimated model is not flexible and it doesn't fit nor predict reasonable results.
First preliminary analysis, Forecast combination, is motivated by one of the most recent studies on Probability of Default models. It demonstrates method of weighting predictions from multiple linear models and shows that combined predictions are under multiple settings better than performance of the best individual models. Part of this study is replicated in this thesis for Czech data and it doesn't support the original hypothesis from the paper. This approach however didn't use any lagged variables, therefore it can't be seen as a candidate model to outperform the original Arimax model.
On the basics of the Forecast combination is built the first competing model framework, Frequentist Model Averaging (FMA). This approach combines model coefficients from multiple linear models to construct final model coefficients. FMA models yields different results in different segments. In the Corporate segment, combined FMA models fit better than the best individual candidate models for given models across all FMA settings. The same doesn't hold for other two sectors, where some of the models combinations fit worse than their best individual models. In both segments however the best performing versions still outperform their best individual models. Performance of FMA models was the best in Corporate sector, where most of the FMA model versions fit better than the Arimax model on the full sample. The best FMA model outperforms the Arimax model in all periods. In the oldest testing period around year 2008, FMA model is the best among all models in Corporate sector. FMA performed decently in Housing sector as well. Most of the models fit better on the full sample than Arimax as well. Performance in testing period around year 2016 isn't that much better than in case of Arimax, however the best FMA model performs better in all testing periods. Poor performance yield the FMA models in the Consumer sector, where even though the models fit better on the full sample, they can't match with the Arimax model in testing periods.
The main competing models were Bayesian Model Averaging models (BMA), which is the framework currently used in ECB. BMA performs better in all the segments. In Corporate segment BMA fit better than Arimax model on the full sample. In testing periods around years 2008 and 2016 BMA models predicted better than Arimax model, but in 2008 worse than FMA. However only BMA model was able around year 2012 to predict decline of PD and the end of the crisis high PDs. BMA model performed particularly well in Housing sector, where the best BMA model performed splendidly on full sample and

in all testing periods as well. No other model, including Arimax, could match with BMA in Housing sector. The best BMA model produced decent results in Consumer segment as well, beating Arimax model in most of the testing periods.

Last discussed method were Neural Networks. Neural Network from all the models suffered the most from short time series and often were overfitted. Neural Networks fit better in all segments in the full sample, they often performed poorly in the testing samples. In the corporate sector Neural Networks were not able to produce dynamic and flexible predictions. In Consumer sector on the other hand networks predicted values far from the actual values. The best performance Neural Networks showed in Housing sector, however even there BMA and FMA methods were more successful in predictions.

Original CNB model doesn't use unemployment, which is often used in relevant literature about credit risk models. It turned out that unemployment and its lags are relevant predictors in all three segments, at most in housing and consumer segment. In these sectors is unemployment (mostly its current value and the 3rd lag, in the housing sector the 4th lag as well) appearing in several best individual models as well as it has very high posterior inclusion probabilities in BMA models, very often over 90%. Unemployment is a bit less relevant in the corporate sector according to BMA posterior probabilities, however it still appears in the best individual models.

Overall BMA models showed the best performance and flexibility, followed by FMA. Both are suitable candidates to serve better than current Arimax model. Neural Networks turned out not to be suitable method for this particular context.

This thesis serves as a comparison of methods. Further extensions of the work could be in more detailed work on one particular method to develop one final model.

Bibliography

Baboucek & Jancar (2005): "A var analysis of the effects of macroeconomic shocks to the quality of the aggregate loan portfolio of the czech banking sector." *CNB Working Paper Series* **(2005/1)**.

Blaschke, Jones, Majnoni, & Peria (2001): "Stress testing of financial systems: An overview of issues, methodologies, and fsap experiences." *IMF Working Paper* .

Buckland, Burnham, & Augustin (1997): "Model selection: An integral part of inference." *Biometrics* **53(2)**: pp. 603–618.

Cihak & Hermanek (2005): "Stress testing the czech banking system: Where are we? where are we going?" *CNB Research and Policy Notes* **(2005/2)**.

Dees, Henry, & Martin (2017): "Stamp€: Stress-test analytics for macroprudential purposes in the euro area." *European Central Bank* .

Deutsche Bundesbank (2015): "Macro stress tests — technical documentation." Deutsche Bundesbank.

Gersl, Jakubik, Konecny, & Seidler (2012): "Dynamic stress testing: The framework for testing banking sector resilence used by the czech national bank." *CNB Working Paper Series* **(2012/11)**.

Henry & Kok (2013): "A macro stress testing framework for assessing systemic risks in the banking sector." *European Central Bank* .

Hjort & Claeskens (2003): "Frequentist model average estimators." *Journal of the American Statistical Associaction* **98**: pp. 879–899.

Jakubik & Schmieder (2008): "Stress testing credit risk: Comparison of the czech republic and germany." *Financial Stability Institute* .

Moral-Benito (2015): "Model averaging in economics: An overview." *Journal of Economic Surveys* **29(1)**: pp. 46–75.

Papadopoulos, Papadopoulos, & Sager (2016): "Credit risk stress testing for eu15 banks: a model combination approach." *Bank of Greece: Working Paper* .

Rumelhart, Hinton, & Williams (1986): "Learning representations by back-propagating errors." *Nature* **323**: pp. 533–536.

Sala-I-Martin, Doppelhofer, & Miller (2004): "Determinants of long-term growth: A bayesian averaging of classical estimates (bace) approach." *American Economic Review* **94(4)**: pp. 813–835.

Wang, Zhang, & Zou (2009): "Frequentist model averaging estimation: A review." *Jrl Syst Sci & Complexity 2009* pp. 732–748.

Willmott & Matsuura (2005): "Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance." *Clim Res* **30(1)**.

CNB Financial Stability Report 2006 (2007): Czech National Bank.

CNB Financial Stability Report 2007 (2008): Czech National Bank.

CNB Financial Stability Report 2008/2009 (2009): Czech National Bank.

CNB Financial Stability Report 2009/2010 (2010): Czech National Bank.

CNB Financial Stability Report 2010/2011 (2011): Czech National Bank.

CNB Financial Stability Report 2011/2012 (2012): Czech National Bank.

CNB Financial Stability Report 2012/2013 (2013): Czech National Bank.

CNB Financial Stability Report 2013/2014 (2014): Czech National Bank.

CNB Financial Stability Report 2014/2015 (2015): Czech National Bank.

CNB Financial Stability Report 2015/2016 (2016): Czech National Bank.

CNB Financial Stability Report 2016/2017 (2017): Czech National Bank.

CNB Financial Stability Report 2017/2018 (2018): Czech National Bank.

Appendix A

Electronic Sources

Here is a hyperlink to website from which you can obtain data and results presented in this thesis:

https://drive.google.com/drive/folders/1bUjwIwl-w6AgSflYCVNhE5V9QE21DpDP?usp=sharing