Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

MASTER THESIS

Probability of default modelling using macroeconomic factors

Author: **Bc. Monika Zsigraiová**

Supervisor: **PhDr. Jakub Seidler, Ph.D.**

Academic Year: **2013/2014**

Declaration of Authorship

The author hereby declares that she 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, July 29, 2014

Signature

Acknowledgments

I would like to thank to my supervisor, PhDr. Jakub Seidler, Ph.D., for his valuable comments and suggestions. Without his guidance and help this thesis would not have materialized. I also thanks to Veronika Dovhunová and Martin Hnath for their great support.

Abstract

The thesis evaluates relationship between probability of default of non-financial corporations and households and evolution of macroeconomic environment. This work contributes to the literature of credit risk proving importance of macroeconomic variables in determining the PDs both on aggregate level and for sector of non-financial corporations and sector of households in the Czech Republic. Evaluation of an impact of the recent financial crisis on the PDs are done by employing latent factor model and FAVAR model on monthly data of non-performing loans and other macroeconomic variables covering the period 01/2002–06/2013. Finally, an ability to forecast and fit the data of FAVAR model and one factor latent model are compared. The comparison indicates that latent factor model should be more appropriate than FAVAR model.

JEL Classification	G21, G28, G33
Keywords	Credit risk, Economic cycle, One-factor model, VAR

Author's e-mail	monika.zsigrai@gmail.com
Supervisor's e-mail	Seidler@email.cz

Abstrakt

Práce posuzuje vztah mezi pravděpodobností selhání nefinančních podniků a domácností a vývojem makroekonomického prostředí. Práce přispívá k literatuře úvěrového rizika potvrzením důležitosti makroekonomických veličin při modelování pravděpodobnosti selhání, a to jak z agregovaného pohledu, tak zvlášť pro sektor nefinančních podniků a domácností v České republice. Vyhodnocení dopadů nedávné finanční krize na pravděpodobnost selhání je provedeno za pomoci Latent factor modelu a FAVAR modelu na měsíčních datech úvěrů v selhání a ostatních makroekonomických ukazatelů pokrývající období 01/2002–06/2013. Na závěr práce vzájemně porovnává schopnost předpovědi skutečných hodnot FAVAR modelu a latent factor modelu. Srovnání naznačuje, že Latent factor model je pro odhady pravděpodobnosti selhání vhodnější než FAVAR model.

Klasifikace JEL	G21, G28, G33
Klíčová slova	Úvěrové riziko, Ekonomický cyklus, One-factor model, VAR

E-mail autora	`monika.zsigrai@gmail.com`
E-mail vedoucího práce	`Seidler@email.cz`

Contents

List of Tables

List of Figures

Acronyms

AR Autoregressive Model

ARCH Autoregressive Conditional Heteroskedasticity

ARIMA Autoregressive Integrated Moving Average Model

CPI Consumer Price Index

CNB Czech National Bank

EAD Exposure at Default

FAVAR Factor-Augmented Vector Autoregression

GDP Gross Domestic Product

IRB Internal-Rated Based

LGD Loss Given Default

MAPE Mean Absolute Percantage Error

MFE Mean Forecast Error

NPL Non-Performing Loans

PC Principal Component

PD Probability of Default

PPI Production Price Index

PRIBOR Prague InterBank Offered Rate

RMSE Root Mean Square Error

SUR Seemingly Unrelated Regression

Var Variance

VAR Vector Autoregression

Master Thesis Proposal

Author	Bc. Monika Zsigraiová
Supervisor	PhDr. Jakub Seidler, Ph.D.
Proposed topic	Probability of default modelling using macroeconomic factors

Topic characteristics Probability of default (PD) is a key credit risk parameter for estimating credit losses. It has become even more important in the last decade, as it serves as an important input factor for determining the minimum capital requirements for banks using IRB (internal-rated based) approach. As ability of debtors to pay back their debts is influenced by economic cycle, macroeconomic factors play an important role in models estimating PD of particular debtor. Especially for a small open economy any macroeconomic shock might be even more important and PD modelling should be linked to the macroeconomic developments. Thus, this thesis will examine several models of PD of Czech economy using monthly data for inflation, interest rate, unemployment, stock index, GDP growth (or its proxy derived from quarterly data) and other determinants (proposed for example in Chan-Lau (2006)). The main source of the data is ARAD database of Czech National Bank. Furthermore, the thesis will model and compare PD for individual sectors (households, corporate sector, government, financial sector and non-residents) as well. The main models used in the analysis will be Logit and Probit models and latent factor model applied for example by Rösch (2003) which is based on Merton (1974).
Moreover, during the examined period, the global markets including the Czech credit market were hit by the financial crisis. Thus, an impact of the crisis will be evaluated using appropriate dummy variables in each of the mentioned models. Finally, resilience of Czech credit market will be assessed according

to the responses to changes in macroeconomic variables, and to make models more dynamic, VAR methodology will be used to test persistency of PD.

Hypotheses

1. Examined macroeconomic factors (GDP growth, interest rate, unemployment, etc.) have significantly influenced PD in all of the models.

2. The latent factor model can predict the PD better than Logit and Probit models.

3. The global financial crisis had an impact on PD in Czech Republic.

4. Czech credit market is resilient to shocks in the macroeconomic factors.

Methodology The most common way how to econometrically model probabilities is to employ Probit and Logit models which will be used in the first part of empirical section of the thesis. Jakubík (2006) argues that these types of models are not appropriate when modelling PD of whole economy so he proposes to use one-factor model (from class of latent factor models) based on Merton´s structural model which relies on the return of assets. The return of assets of a subject is regressed on the macroeconomic variables, and probability of default is extracted using log-likelihood function. To test hypothesis 2, the prediction of the models and real data will be compared and the best model will be identified.
Next part of the thesis will test the persistence of PD in time to determine to what extent history is relevant for current PD. Thus, the VAR methodology with lagged PD and other explanatory variables will be employed as proposed by Simons and Rolwes (2009). Finally, VAR methodology (impulse responses respectively) will be also used to test the resilience of Czech economy and its credit market - the best model of PD will predict sufficient number of future values and a simulation of shocks in PD and in the most important macroeconomic variables will be employed to determine an influence on GDP and unemployment.

Outline

1. Introduction
2. Literature Overview
3. Methodology and Data
   (a) Logit and Probit Models
   (b) Latent Factor Models
   (c) Persistence and Resilience
4. Empirical results and its discussion
5. Conclusion

Core bibliography

1. Câmara, A., I. Popova, & B. Simkins(2012): "A comparative study of the probability of default for global financial firms. Journal of Banking & Finance." *European Economic Review* **36(3)**: pp. 717–732.
2. Chan-Lau, J. (2006): "Fundamentals-based estimation of default probabilities: A survey."
3. Hamerle, A., A. Dartsch, R. Jobst, & K. Plank (2011): "Integrating macroeconomic risk factors into credit portfolio models." *The Journal of Risk Model Validation* **(5)**: pp. 3–24.
4. Jakubík, P. (2006): "Does Credit Risk Vary with Economic Cycles? The Case of Finland." *Charles University in Prague - IES Working Paper, no. 11* **50(2)**: pp. 285–305.
5. Jakubík, P. (2007): "Macroeconomic Environment and Credit Risk." *Czech Journal of Economics and Finance* **57(1-2)**: pp. 41–59.
6. Karminsky, A., A. Kostrov, & T. Murzenkov (2012): *"Comparison of Default Probability Models: Russian Experience?"* Higher School of Economics Research Paper No. WP BRP, 6.
7. Merton, R. C. (1974): "On the pricing of corporate debt: The risk structure of interest rates." *The Journal of Finance* **29(2)**: pp. 449–470.
8. Koopman, S. J., & A. Lucas (2005): "Business and default cycles for credit risk." *Journal of Applied Econometrics* **20(2)**: pp. 311–323.
9. Lowe, P. (2002): "Credit risk measurement and procyclicality."
10. Pesaran, M. H., & T. Schuermann (2003): "Credit risk and macroeconomic dynamics." *Medium Econometrische Toepassingen* **11(1)**: pp. 27–32.
11. Rösch, D., (2003): "Correlations and business cycles of credit risk: Evidence from bankruptcies in Germany." *Financial Markets and Portfolio Management* **17(3)**: pp. 309–331.

12. Simons, D., & F. Rolwes (2009): "Macroeconomic default modeling and stress testing." *International Journal of Central Banking* **5(3)**: pp. 177–204.

Author

Supervisor

Chapter 1

Introduction

Probability of default (PD) is a key credit risk parameter for estimating credit losses. It has become even more important in the last decade, as it serves as an important input factor for determining the minimum capital requirements for banks using IRB (internal-rated based) approach and banks own internal assessment processes. Aggregated probability of default is not only credit risk parameter, but also the most important input of stress testing in the Czech Republic as referred for example by Jakubík & Heřmánek (2006) or Geršl & Seidler (2010).

Prior literature examines the importance of macroeconomic variables in determining probability of default. Koopman & Lucas (2003), for example, reject a significant influence of growth rates on business default. On the other hand, plenty of studies (e.g. Rösch (2005), Marcucci & Quagliariello (2009), or Hamerle *et al.* (2011)) recommend an inclusion of macroeconomic variables into estimated equations to get better results. Thus, the hypothesis is to test an ability of macroeconomic variables to determine probability of default in the Czech Republic. Moreover, the hypothesis of influence of the macroeconomic variables on sectors of non-financial corporations, households, and non-residents separately is tested as well.

Credit risk methodology is usually based on availability of data. In this study, the monthly data of aggregated non-performing loans and other macroeconomic variables (1/2002–6/2013) provided by CNB are used. With respect to data frequency and their availability, the chosen models are latent factor model (recommended by Basel II) and FAVAR (which should control for collinearity between macroeconomic variables and enables to employ all the macroeconomic variables). Jakubík (2006) claims that the vector autoregression models do not perform as good as the latent factor models due to a possible non-linearity. Therefore another hypothesis that latent factor model better fit the data and better forecast actual probability of default than the FAVAR model is tested. Furthermore, data contains a period of financial crisis and there might be a structural break in the probability of default in September 2008 (Figure 4.1 and 4.2). Hence, the thesis examines another hypothesis that the Czech probability of default is influenced by the crisis employing dummy variables.

It is contributed to the literature of credit risk proving the macroeconomic variables important in determining the PDs of all the sectors in the Czech Republic. Moreover, analysis indicates an influence of the crisis to be significant when all sectors together, the sector of non-financial corporations, and the sector of non-residents are modelled. The sector of households was not proven to be seriously hit by the crisis. Furthermore, the latent factor model could be considered as a slightly better one according to the forecasting measures employed. On the other hand, FAVAR model, firstly used for a Czech credit risk data, enables to include more variables which are mutually correlated and the resulting forecasts do not differ so much. Finally, the monthly data might also provide an advantage of more observations to other studies (e.g. Jakubík & Schmieder (2008)).

The rest is structured as follows. Chapter 2 summarizes recent literature when Chapter 3 lists and describes models of probability of default concerning the models using macroeconomic variables. Chapter 4 describes data and provides definitions of probability of default as well as definitions of macroeconomic determinants. Chapter 5 describes results of both methods for all the sectors and Chapter 6 concludes a discussed the results.

Chapter 2

Literature Overview

Probability of default (PD) is connected to the borrower grade and it is defined by the Basel committee. PD must be internally estimated by banks when using foundation internal rating based approach. Values of Loss given default (LGD) and Exposure at default (EAD) are given by supervisory rules in this case. Those estimations and maturity attached to the exposure determine the level of risk-weighted assets and regulatory capital which every bank must keep according to the Basel Committee. Basel Committee also permits for advanced internal rating based approach when LGD and EAD are estimated as well as the PD, so the PD has to be estimated in both cases. (Basel Committee on Banking Supervision 2001)

The Basel Committee on Banking Supervision (2004) further specifies that the estimations of PD and possibly other determinants (LGD, EAD) should be key factors in credit approval process, risk management, internal capital allocations and also corporate governance of regulated bank. On the other hand, the Basel Committee on Banking Supervision (2004) also claims that estimates for IRB purposes and internal estimates probably will not be the same. Furthermore, it is here emphasized that every regulated bank should perform sound stress testing process including scenarios with unexpected negative shocks and it should evaluate its effect on capital requirements.

Following Basel Committee on Banking Supervision (2001), Bank for International Settlements also suggested to discuss measurement of credit risk, especially its connection to business cycles in Lowe (2002). The paper suggests more research in the area of macroeconomic conditions and credit risk because of difficult determination of financial imbalances and it also discussed changes in capital requirement in different macroeconomic conditions. He recommends an inclusion of the macroeconomic variables into a credit risk modelling and regulatory capital assessment even if the imbalances cannot be quantified. Paper further points out that the level of the capital should not be decreased during the financial imbalances. Lowe (2002) appeal to supervisors and markets to be able to ensure, that those financial imbalances will not make the macroeconomic fluctuations larger.

Basel Committee on Banking Supervision (2001) also requires using a time series longer than 5 years for estimation of the PD. Stein (2006) argues that large datasets might have problem with correlation among firms in the sample. Therefore, he claims that more observations do not mean better performance of the model necessarily and confidence intervals might be larger. Narrowing the intervals was also object of Schuermann & Hanson (2004) as they compared different estimation methods of cohort and density. They found out that intervals are very tight and it is very hard to distinguish the rating levels on investment grade levels. However, on the speculative level, differentiation between levels is quite clean according to them. Schuermann & Hanson (2004) also indicate that the differentiation is also conditioned by state in which the business cycle is. They suggest that it is easier to differentiate between the grades in case of depression than in case of expansion.

Rösch (2005) refers that, following Basel II, there are two main types of credit ratings - Through the cycle and Point in time. He explains that Through the cycle models are used primary by credit rating agencies as they should reflect long term probabilities of default that are not dependent on business cycle. On the other hand, he also claims that Point in time models concern about present moment or future pre-specified horizon and conditions which are influencing probabilities. Rösch (2005) concludes that the Point in time model performs lower correlations between assets, so the Point in time model might be more appropriate for probability of default forecasting. The Point in time models are also recommended by Hamerle *et al.* (2003) or Hamerle *et al.* (2011). Hamerle *et al.* (2011) emphasize that banks usually employ Through the cycle model, but he shows that adding macroeconomic variables into model (i.e. makes it Point in time) significantly improves the model.

Because of indicated Point in time models appropriateness, there are plenty of studies that connect probability of default with business cycles for example Koopman & Lucas (2003), Hamerle *et al.* (2004), Pederzoli & Torricelli (2005), or Marcucci & Quagliariello (2009). First, Koopman & Lucas (2003) reject the hypothesis that there is a strong influence of the growth rates on business default, although a correlation between GDP growth and defaults exists. Other studies rather support the importance of business cycle presence in credit risk models. Hamerle *et al.* (2004) claims that inclusion of variables correlated with business cycle significantly improves the forecasting power of the model. Marcucci & Quagliariello (2009) also examined differences between expansion and recession taking into account riskiness of banks. They conclude that all the banks are affected by business cycle and that credit risk is higher during a recession.

To test significance of specific macroeconomic variables several models was built. One of them is model used by Virolainen (2004) and followed by Fiori *et al.* (2009) and Simons & Rolwes (2009). Virolainen (2004) is using logistic form estimated by SUR. He does not model whole economy but corporate sector divided to six industries. Significance between probability of default and macroeconomic variables was found by him for GDP growth, interest rate, and indebtness of the corporate sector. Fiori *et al.* (2009) follow Virolainen (2004), but they do not test specific macroeconomic variables. They create latent factors by Principal Component Analysis and those factors were proven by them to be significant determinant of probability of default as well. Finally, Simons & Rolwes (2009) test also oil price besides other macroeconomic variables. They conclude that the oil price is significant determinant of Dutch corporate sector as well as the GDP growth, interest rate, and exchange rate.

Another frequently used methodology of probability default estimation is VAR as it adds more dynamic into a model. It is used for example by Jakubík (2006), Marcucci & Quagliariello (2008), Alessandri *et al.* (2009), or Hamerle *et al.* (2011). Marcucci & Quagliariello (2008) confirm that probability of default follow cyclical pattern. Alessandri *et al.* (2009) and Hamerle *et al.* (2011) also add macroeconomic variables (GDP growth, unemployment, inflation, interest rates, Dow Jones Index, etc.) into the regressions which are used to probability of default forecasting. On the other hand, Jakubík (2006) does not consider the VAR model to be best model for probability of default modelling due to the non-linearities and he recommends to use Merton type models.

Merton type models are based on Merton (1974) which is built on debt valuation using option pricing theory of Black & Scholes (1973). Those types of models are often connected with unobservable factor and they are called latent factor models. The latent factor modes are employed for example by Hamerle *et al.* (2003), Jakubík (2006), Jakubík (2007), Jakubík & Schmieder (2008), Ali & Daly (2010) and Hamerle *et al.* (2011). Hamerle *et al.* (2003) emphasize that this model should be used in case appropriate proxies for asset returns are available. Basic one latent factor model with macroeconomic variables is employed by Jakubík (2006) for probability of default of Finish economy, Jakubík (2007) modelling probability of default of Czech economy, Jakubík & Schmieder (2008) for probability of default of Czech and German sector of non-financial corporations and households, and by Ali & Daly (2010) comparing US and Australian probability of default responses to macroeconomic shocks. All of the studies confirm significance of macroeconomic variables in probability of default modelling. Finally, Hamerle *et al.* (2011) add more factors and estimates multifactor latent model with unemployment, Dow Jones Index and index of industrial production for forecasting of probability of default.

Jakubík (2007) also makes the one latent factor model applicable to the Czech non-performing loans (NPL) data collected by Czech National Bank. In his paper, he concludes that all the macroeconomic variables examined (GDP growth, inflation, interest rate, and indebtness) should significantly influence loans probability of default in the Czech economy. The Czech sector of households and non-financial corporation is then analyzed in Jakubík & Schmieder (2008) with similar conclusions about significance of macroeconomic determinants. Those probability of default models are further purposed to use as a credit risk input in stress testing of central bank as for example in Jakubík & Heřmánek (2006) or Geršl & Seidler (2010).

Chapter 3

Probability of Default Models

Usually, the chosen model from great variety of models of Probability of Default depends on the data that are available to the researcher. Chan-Lau (2006) divides the models according to the data nature into two main groups - market models and fundamental based models where market models of credit risk are based on security prices and fundamental based models takes into account rating information, systematic market and economic factors and accounting. Thus, he classifies models into classes of:

- Accounting based (or credit scoring) models - here belongs for example Altman Z-score or Moody´s KMV,
- Rating based models - used for instance in Schuermann & Hanson (2004) or Rösch (2005),
- Macroeconomic models - based on macroeconomic variables - for example Virolainen (2004) or Jakubík (2006),
- Hybrid models - combination of models mentioned above.

Macroeconomic models are the main concern of the thesis. Chan-Lau (2006) further divides macroeconomic based model into two groups - models with exogenous and endogenous economic factors. He emphasizes that both models are based on the fact that during the crisis more defaults can be observed and vice versa.

3.1 Endogenous economic factors models

The most common endogenous model is VAR model - the VAR (p) (similarly to Chan-Lau (2006))is defined as:

$$V_t = c + \sum_{i=1}^{p} a_{t-i} V_{t-i} + e_t \tag{3.1}$$

where Vt is endogenous variable (vector of dimension $k \times 1$) containing PD as well as macroeconomic determinants, c is constant (dimension $k \times 1$), e_t represents shocks (dimension $k \times 1$), k is number of endogenous variables, p is number of lags and a_{t-i} are estimated parameters for all $i = 1, \ldots, p$. This model can be further examined by impulse response analysis which was used for example in Alessandri *et al.* (2009) or Hamerle *et al.* (2011).

3.2 Exogenous economic factors models

The probability of default at time t (PD_t) is usually described by some function of macroeconomic variable. Chan-Lau (2006) uses further definition:

$$PD_t = f(y_t)$$
$$y_t = g(x_t, e_t)$$

where the macroeconomic variable y_t is a function of given macroeconomic determinants x_t and shock e_t. One of these groups of models is model used by Virolainen (2004). He examines PD of several sectors of economy. Thus, he specifies the probability function of macroeconomic variables in logistic form:

$$p_{j,t} = \frac{1}{1 + exp(y_{j,t})} \tag{3.2}$$

where $p_{j,t}$ is probability of default of industry j and $y_{j,t}$ is industry specific macroeconomic index. Thus, logit transformation results in:

$$L(p_{j,t}) = \frac{1 - p_{j,t}}{p_{j,t}} = y_{j,t} \tag{3.3}$$

Furthermore, industry specific macroeconomic index is treated bz him in the following way:

$$y_{j,t} = b_{j,0} + b_{j,1} X_{1,t} + ... + b_{j,n} X_{n,t} + \nu_{j,t} \tag{3.4}$$

where $X_{i,t}$ is i-th explanatory macroeconomic variable at time t, $b_{j,i}$ is estimated coefficient of i-th variable in sector j, $\nu_{j,t}$ represents a shock in i-th industry at time t and $i = 1, \ldots, n$. $X_{i,t}$ is assumed to follow autoregressive process of order p (AR(p)).

3.2.1 Latent Factor Models

One Factor Model

Another class of exogenous models – the latent factor models – are used to explain effects of explanatory variables (or factors) which are not observable. Basic framework for using factor models in PD modelling is stated by Basel II and summarized for example in Hamerle *et al.* (2003). The returns on asset i at time t are defined here as:

$$R_{i,t} = Y_{i,t} + y_{i,t-1} \tag{3.5}$$

for $i = 1, ..., N$ and $t = 1, ..., T$. Moreover, it is assumed that random variable of returns $R_{i,t}$ has normal distribution with mean $\mu_{i,t}$ and variance σ_i^2 ($N(\mu_{i,t}, \sigma_i^2)$). $Y_{i,t}$ is random variable of value of asset i at time t and $y_{i,t-1}$ is its realization in previous period of time.

$$E(R_{i,t}|y_{i,t-1}) = \mu_{i,t} \tag{3.6}$$

$$Var(R_{i,t}|y_{i,t-1}) = \sigma_i^2 \tag{3.7}$$

This return is modelled by them using two types of explanatory variables. First, the F_t is factor which is common for all the market at time t (systematic factor). The second variable is $U_{i,t}$ which is assumed to represent idiosyncratic part – the individual influence on i-th asset at time t. Moreover, both of the variables fulfil the assumption of their standard normal distribution and they are serially independent and independent on each other according to them.

$$R_{i,t} = \mu_{i,t} + bF_i + \omega U_{i,t} \tag{3.8}$$

for $i = 1, ..., N$ and $t = 1, ..., T$ and $F_t \backsim N(0; 1)$, $U_{i,t} \backsim N(0; 1)$. Next, Hamerle *et al.* (2003) distinguish between two default probabilities under the Basel II framework – conditional and unconditional on previous factor realization. The probability of default is defined by them as a probability that the value of asset i at time t falls below given threshold $c_{i,t}$. Formally:

$$\begin{aligned} PD_{i,t} &= P(Y_{i,t} < c_{i,t}) = \\ &= P(R_{i,t} < c_{i,t} - y_{i,t-1}) = \\ &= P(\frac{R_{it} - \mu_{it}}{\sigma_i^2}) < \frac{c_{it} - y_{i,t-1} - \mu_{it}}{\sigma_i^2}) = \\ &= F_N(\frac{c_{it} - y_{i,t-1} - \mu_{it}}{\sigma_i^2}) \end{aligned}$$

where F_N is cumulative standard normal distribution function. They claim that conditional probability of default is probability dependent on realization f_t of the systematic random variable F_t. Hence,

$$\begin{aligned} PD_{i,t} &= P(Y_{i,t} < c_{i,t}|f_t) = \\ &= P(R_{i,t} < c_{i,t} - y_{i,t-1}|f_t) = \\ &= P(\frac{R_{it} - \mu_{it}}{\sigma_i^2}) < \frac{c_{it} - y_{i,t-1} - \mu_{it}}{\sigma_i^2}|f_t) \end{aligned}$$

Furthermore, Hamerle *et al.* (2003) specify factor model containing macroeconomic variables in form of:

$$R_{i,t} = \beta_{0,i} + \beta_i^T Z_t + \omega U_{i,t} \tag{3.9}$$

for $i = 1, ..., N$ and $t = 1, ..., T$. Hence, the return of subject i at time t $R_{i,t}$ is explained by matrix of macroeconomic variables (GDP, inflation, unemployment,...) at time t (Z_t) and subject specific idiosyncratic part ($U_{i,t}$). They also emphasize that the expected return and the expected returns conditional on realization z_{t-1} of Z_t should be time independent – for every t must hold:

$$E(Z_t) = E(Z) \tag{3.10}$$

$$Var(Z_t) = Var(Z) \tag{3.11}$$

then:

$$E(R_{i,t}) = \beta_{0,i} + \beta_i^T Z \tag{3.12}$$

$$E(R_{i,t}|z_{t-1}) = \beta_{0,i} + \beta_i^T Z \tag{3.13}$$

Because returns cannot be observed sometimes, they model probability of default by latent factor model. Firstly, Hamerle *et al.* (2003) define variable Y^* equal to 1 (if default if i-th subject occurs at time t) and 0 (otherwise). Thus, if the realization of systematic part f_t and factors of i-th subject at time t w_{it} are given, the conditional probability of default is defined by them as:

$$\begin{aligned} PD_{i,t}(w_{i,t}, f_t) &= P(Y^*_{i,t} = 1|w_{i,t}, f_t) = \\ &= P(Y_{i,t} < c_{i,t}|w_{i,t}, f_t) = \\ &= P(U_{i,t} < \frac{c_{i,t} - y_{i,t-1} - \beta_{0,i} - \delta' w_{i,t} - bf_t}{\omega}|w_{i,t}, f_t) = \\ &= F(\beta^*_{0,i} + \delta^{*'} w_{i,t} - bf^*_t|w_{i,t}, f_t) \end{aligned}$$

where $F(.)$ is cumulative distribution function of distribution of idiosyncratic term $U_{i,t}$,

$$\beta^*_{0,i} = \frac{c_{i,t} - y_{i,t-1} - \beta_{0,i}}{\omega} \tag{3.14}$$

$$\delta^{*'} = \frac{-\delta'}{\omega} \tag{3.15}$$

$$b^* = \frac{b}{\omega} \tag{3.16}$$

Hence, they suggest that in term $\beta^*_{0,i}$, there are hidden latent factors which cannot be observed – threshold level, asset value and subject specific intercept. Furthermore, probability of default of this model depends on the selection of distribution F_N of idiosyncratic part of returns. Usually, it is assumed standard normal distribution (for example Jakubík (2006)), Hamerle *et al.* (2003) suggest using Logit model. i.e.

$$PD_{i,t}(w_{i,t}, f_t) = F(\beta^*_{0,i} + \delta^{*'} w_{i,t} - bf^*_t) \tag{3.17}$$

$$PD_{i,t}(w_{i,t}, f_t) = \frac{exp(\beta^*_{0,i} + \delta^{*'} w_{i,t} - bf^*_t)}{1 + exp(\beta^*_{0,i} + \delta^{*'} w_{i,t} - bf^*_t)} = L(\beta^*_{0,i} + \delta^{*'} w_{i,t} - bf^*_t) \tag{3.18}$$

To obtain unconditional probability of default, Hamerle *et al.* (2003) further integrate the conditional function over the realizations f_t:

$$PD_{i,t}(w_{i,t}, f_t) = \int_{-\infty}^{+\infty} F(\beta^*_{0,i} + \delta^{*'} w_{i,t} - b f^*_t)\phi(f_t) df_t \qquad (3.19)$$

where $\phi(.)$ is standard normal distribution density function.
Usually, the returns of given assets are represented by following function used for example by Jakubík (2006):

$$R^*_{it} = \sqrt{\rho} F^*_t + \sqrt{1-\rho} U^*_{it} \qquad (3.20)$$

where ρ represents correlation between normalized returns of assets of any two subjects, R^*_{it} is random variable of normalized logarithmic return of i-th subject at time t, F^*_t is random variable of normalized logarithmic return at time t which is not dependent on the i-th subject, and finally, U^*_{it} represents random variable of normalized logarithmic return at time t which is linked to the subject i. Moreover, Jakubík (2006) assumeds that F^*_t and U^*_{it} have standard normal distribution and all the random variables are serially independent and the distribution of R^*_{it} is also normal with expected value of zero and its variance equals to one.

$$F^*_t \backsim N(0;1)$$
$$U^*_{it} \backsim N(0;1)$$
$$R^*_{it} \backsim N(0;1)$$

Jakubík (2006) further states that the default takes place when returns drop below certain threshold T. This threshold is also modelled by him using j macroeconomic variables. Hence, he express the probability of default in the following structure:

$$PD_{i,t} = P(\sqrt{\rho} F^*_t + \sqrt{1-\rho} U^*_{it} < \beta_0 + \sum_{j=1}^{N} \beta_j x_{jt} = F_N(\beta_0 + \sum_{j=1}^{N} \beta_j x_{jt}) \qquad (3.21)$$

For realization f^* of logarithmic returns F^*, the conditional default probability is expressed by Jakubík (2006) as:

$$PD_i(f_t^*) = P(U_{it}^* < \frac{\beta_0 + \sum_{j=1}^{N} \beta_j x_{jt} - \sqrt{\rho} f_t^*}{\sqrt{1-\rho}}) = F_N(\sqrt{\rho} f_t^* \sqrt{1-\rho}) \tag{3.22}$$

Next, the law of large number is used by him to prove that the probability of default of i-th subject is the same as the probability of whole portfolio. According to the data which contains defaults represented by whole sector/economy, Jakubík (2006) refers it holds:

$$P(PD(f_t^*) = PD_i(f_t^*)|F_t^* = f_t^*) = 1 \tag{3.23}$$

Furthermore, Jakubík (2006) states that the unconditional probability is expressed as:

$$PD = \int_{-\infty}^{+\infty} PD(f_t^*)\phi(f_t^*)df_t^* \tag{3.24}$$

and model probability of default employing binomial distribution. He supposes that if the condition default probability $PD(f_t^*)$ and number of subjects N_t in the sector/economy at time t are given, the number of defaults at time t $D_t(f_t^*)$ has binomial distribution. Further, he states that conditional (on f_t^*) and unconditional probabilities that realization d_t of D_t takes place are following:

$$P(D_t = d_t|F_t^* = f_t^*) = \binom{n_t}{d_t} PD(f_t^*)^{d_t}(1 - PD(f_t^*))^{n_t - d_t} \tag{3.25}$$

$$P(D_t = d_t) = \int_{-\infty}^{+\infty} \binom{n_t}{d_t} PD(f_t^*)^{d_t}(1 - PD(f_t^*))^{n_t - d_t}\phi(f_t^*)df_t^* \tag{3.26}$$

where n_t is realization of random variable of number of subjects in given sector N_t.

Multi Factor Model

Not only one factor model is employed in analysis of probability of default. There are also latent multi factor models used for example by Hamerle *et al.* (2011). They use observable subject specific as well as macroeconomic variables and latent factor variables to model returns or credit quality when one period lag is assumed:

$$R_{it} = \beta_0 + \beta^T x_{it-1} + \gamma^T z_{t-1} + S_{it} \tag{3.27}$$

where R_{it} is return of subject i at time t, x_{it-1} represents subject specific variables at time $t-1$, z_{t-1} denotes level of macroeconomic variables at time $t-1$ and S_{it} is unobservable part of the model which is modelled by risk sector factor and idiosyncratic factor:

$$S_{it} = \sqrt{\rho_j} f_{jt} + \sqrt{1-\rho_j} U_{it} \tag{3.28}$$

where f_{jt} belongs to the vector $f_t = (f_{1t}, \ldots, f_{Jt})$. Each factor f_{jt} denotes the risk in given sector j at time t. Secondly, U_{jt} is idiosyncratic term of subject i at time t and j is correlation of returns within the same sector.
To estimate the default of subject i, the threshold method to determine the default is also used by Hamerle *et al.* (2011) to model conditional default probability. Then, the maximum likelihood method could be used to estimate the parameters of model.

3.2.2 Hazard Model

Another class of models which use latent factor determinant is the class of hazard models. One of the hazard models is described for example in Chava *et al.* (2011). They define default density (or hazard rated) function as:

$$\lambda_{ij}(t) = Y_i exp(X_{ij}(t)\beta) \tag{3.29}$$

where $\lambda_{ij}(t)$ is probability density function of j-th firm in i-th sector at time t, $X_{ij}(t)$ is vector of macroeconomic as well as firm specific variables of j-th firm in i-th sector at time t, β is vector of parameters to be estimated, and Y_i latent non-negative random factor common for whole i-th sector. Moreover, they claim that the latent factor has assumed distribution which is updated during the time.

3.2.3 FAVAR

Both frequently used in PD modelling, the latent factor model and vector autoregression (VAR), could be put together using factor augmented vector autoregression (FAVAR). FAVAR is often used to determine an effect of macroeconomic variables on interest rates. The plenty of macroeconomic and mutually correlated variables are used to explain monetary transmission mechanism for example in Gupta *et al.* (2010) or Fernald *et al.* (2013).

If the macroeconomic variables are employed in PD modelling, it might be not a good choice to use VAR for several reasons - for example, Fernald *et al.* (2013) claim that the estimation could be infeasible without history of data long enough. Employing of many variables might cause instability in VAR results. This is the reason why the following analysis of PD uses the FAVAR methodology instead of VAR.

As in the most papers using FAVAR, the methodology is based on Bernanke *et al.* (2004). They assume a time series Y_t of dimension $M \times 1$ of observable economic variables and time series of unobservable factor F_t of dimension $K \times 1$. Moreover, they assume that:

$$\begin{bmatrix} F_t \\ Y_t \end{bmatrix} = \Phi(L) \begin{bmatrix} F_{t-1} \\ Y_{t-1} \end{bmatrix} + \nu_t$$

where $\Phi(L)$ is a lag operator of finite order d and ν_t is error term with zero mean and variance Q. They also notes that if the all coefficients of lag operator which connects lagged values of F_{t-1} to Y_t are not zero, the equation could be considered as the representation of FAVAR. Bernanke *et al.* (2004) further mentione that this equation cannot be estimated directly because the factor is unobservable. Thus, another assumption that the factors is derived by them from the economic time series. The series X_t of dimension $N \times 1$ is defined by them as:

$$X_t^T = \Lambda^f F_t^T + \Lambda^y Y_t^T + e^T \tag{3.30}$$

where X_t is informational time series related to the unobservable factor and observable economic factor, Λ^f is factor loading of dimension $N \times K$, Λ^y is $N \times M$ and e is error term with zero mean and it is assumed that errors are weakly correlated.

To estimate those equations, two step procedure is employed by Bernanke *et al.* (2004). In the first step, principal component analysis of series X_t is done to find some estimate of unobserved factor. In the second step, they use those components in standards procedure of vector autoregression.

Principal Component Analysis (PCA)

Principal component analysis reorganize variables in the way that the most of the variance is explained by the first principal component, the second component explains the second largest share of variance in the data etc. It is done by linear transformation of the data explained for example in Jolliffe (2002) – the time series x of dimension $1 \times p$ is multiplied by vector of p constants α_1 to get linear combination

$$\alpha_1^T x = \alpha_{11}x_1 + \alpha_{12}x_2 + \ldots + \alpha_{1p}x_p \tag{3.31}$$

which maximize the variance

$$var(\alpha_1^T x) = \alpha^T \Sigma \alpha_1 \tag{3.32}$$

where Σ is variance-covariance matrix of x. The optimum condition using restriction that $\alpha_1^T \alpha_1 = 1$ is then

$$(\Sigma - \lambda_1 I_p)\alpha_1 = 0 \tag{3.33}$$

where is λ_1 Lagrange multiplier, I_p is identity matrix of dimension $p \times p$. Therefore, λ_1 is eigenvalue of Σ and α_1 is its eigenvector. Because of employing the restriction, maximizing the variance means to maximize eigenvalue λ_1. Then, the second principal component has the second largest eigenvalue λ_2, etc. up to the λ_k, where k is required/maximum number of components of x. Moreover, all of the principal components of x are not correlated. (Jolliffe 2002)
Finally, there are two methods which can be employed in principal component analysis – the method using correlation matrix or covariance. The method using correlation matrix instead of covariance was chosen because of different scales of used variables.

Vector Autoregression

The next step of Bernanke *et al.* (2004) is to apply vector autoregression, but instead of the macroeconomic variables, the principle components are used. Generally, the VAR is described by above mentioned equation:

$$V_t = c + \sum_{i=1}^{p} a_{t-i} V_{t-i} + e_t \tag{3.34}$$

where V_t is endogenous variable (vector of dimension $k \times 1$) containing PD as well as k principal components, c is constant (dimension $k \times 1$), e_t represents shocks (dimension $k \times 1$), k is number of endogenous variables, p is number of lags and a_{t-i} are estimated parameters for all i= $1, \ldots, p$. Hence for FAVAR model:

$$V_t = [PD_{it}, PC1, \ldots, PCk]^T \tag{3.35}$$

where is PD_{it} it probability of default of sector i at time t and PCk denotes k-th principal component. Moreover, this model should be further examined by impulse response analysis and variance decomposition.

3.2.4 Forecasting Measures

To assess ability of the credit risk models to fit the actual values and forecast, the forecasting measures should be employed. The used forecasting measures following Kennedy (2003) are:

- Mean Absolute Percentage Error (MAPE):
  This measure should return an average of absolute values of percentage difference between actual and forecasted values. From the definition, a problem might rise if the actual value is equal to zero, then the measure is not defined.

$$MAPE = \frac{\sum_{n=t+1}^{T} \left| \frac{actual - forecast}{actual} \right|}{T - t} \tag{3.36}$$

- Mean Forecast Error (MFE):
  The mean forecast error employs average of a difference between actual and forecasted values over forecasting period.

$$MFE = \frac{\sum_{n=t+1}^{T} (actual - forecast)}{T - t} \tag{3.37}$$

- Root Mean Square Error (RMSE):
  RMSE is defined as square root of an average of squared differences between actual and forecasted values. Then, RMSE penalized smaller values more than larger ones.

$$RMSE = \sqrt{\frac{\sum_{n=t+1}^{T}(actual - forecast)^2}{T - t}} \tag{3.38}$$

where *actual* are actual values of PD, $forecast$ are forecasted values, T is end of forecasting period and t is the end of observed period used for estimating the model.

Chapter 4

Employed Variables and Data

The data used to estimate given models comes from ARAD database of Czech National Bank; the frequency of the data is monthly (1/2002–6/2013). Using monthly frequency for estimating models is one of the main contributions of the thesis as other studies used quarterly data only and the monthly data may reveal relationships which are not included in quarterly data because of more observations available. Furthermore, the data can be easily separated to two groups to assess forecasting ability of the estimated models. The data are separated into in-sample data covering period of 1/2002–12/2012 (which are used for analysis purposes) and out-of-sample data covering 1/2013–6/2013 employed to control quality of analysed models. Moreover, the length of data satisfies a condition by Basel Committee that observations should be longer than 5 years when computing PD (Basel Committee on Banking Supervision 2001).

4.1 Probability of Default (PD)

For the purposes of an analysis, Simons & Rolwes (2009), for example, define PD as a ratio between number of defaults and average number of firms. Further, Fungáčová & Jakubík (2012) use a ratio of new bad loans to performing loans in the economy. In line with previous literature (e.g. Fungáčová & Jakubík (2012), Geršl & Seidler (2010)), the probability of default is defined as:

$$PD_{ti} = \frac{NPL_{ti} - NPL_{t-1,i} + rNPL_{t-1,i}}{Total\_Loans_{t-1,i} - NPL_{t-1,i}} \tag{4.1}$$

where PD_{ti} is probability of default in the economy/sector i at time t, NPL_{ti} is amount of non-performing loans in the economy/sector i at time t, $Total\_Loans_{t-1,i}$ is amount of total loans in the economy/sector i at time t, and r is outflow rate of non-performing loans. Hence, the rate r determines how many non-preforming loans were written-off or reclassified to performing loans during a one month.

As most of the variables are directly provided, only the rate of outflow has to be determined. Based on the previous studies Geršl & Seidler (2010), Fungáčová & Jakubík (2012), and Jakubík & Heřmánek (2006), the rate was set at 15%. Because Geršl & Seidler (2010) also claims that the rate is quite volatile, the determined constant is very rough approximation which might overstate and understate the volume of new loans and PD at some points of time. However, it is the best approximation available and it is a standard procedure employed by recent studies.

In line with the fact that the outflow rate is a constant, it has occurred cases (in sectors of households and non-residents) when the rate was too low (especially during the very volatile beginning of the examined period) and the probability of default measure turned to be negative. These points were detected as outliers (less than 4% of observations) and set to the zero. Furthermore, to ensure that the outliers do not significantly affect dependencies between variables, the robustness check on the data containing outliers the follows the analysis.

The PD of all sectors consist aggregate data of sectors of non-financial corporations, financial institutions, government, households, and non-residents. As total loans and non-preforming loans of government and financial institutions covers only marginal amount of total loans (Table 4.1), they are not analysed separately, although they are included in the PD of all sectors. Therefore, only the sectors of non-financial corporations, households and non-residents are modelled. Table 4.1 also indicates that most important sector to PD modelling should be the sector of corporations whose non-performing loans cover more than half of total non-performing loans.

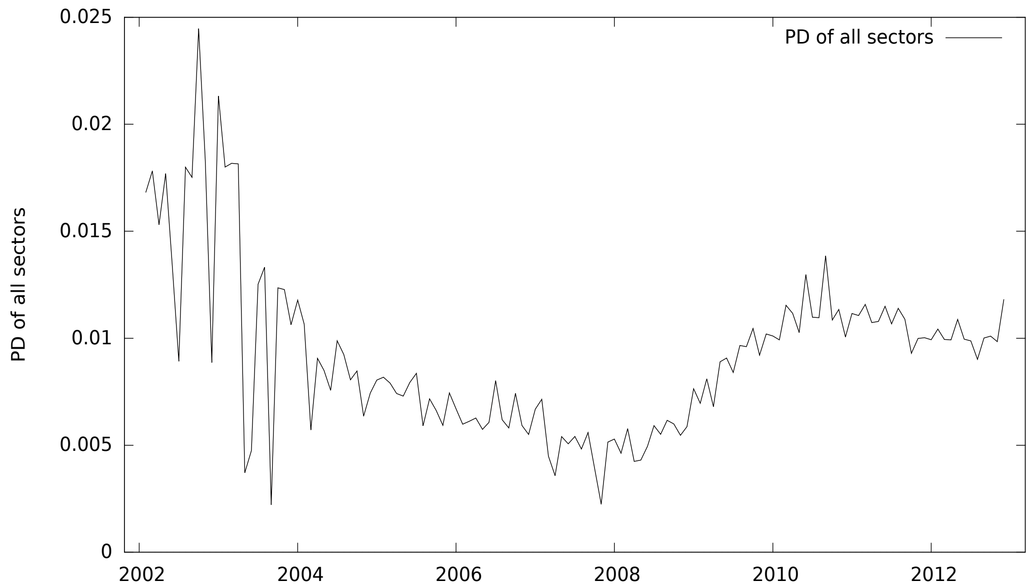
Figure 4.1 shows the evolution of approximated PD ratio of all sectors in time. The PD ratio had declining trend during the years 2002–2007. It decreased by 1.5 percentage points to its low of 0.2% during that period. This trend has changed in the middle of 2008 when the financial crisis hit the Czech Republic. The ratio started to growth, and since 2011, the ratio seems to be constant, slightly fluctuating around 1%. Finally, the ratio seems to be more volatile in the beginning of the examined period when it reached its maximum volatility of 2.5%.

Table 4.1: Average share of Loans/Non-Performing Loans on Total Loans/Non-Performing Loans in the Czech Republic during 01/2002-06/2013

	Share on loans	Share on Loans with Default
Corporations	39.65%	55.26%
Fin.Institutions	6.78%	1.29%
Government	5.76%	0.18%
Households	40.47%	34.24%
Non-residents	6.85%	8.95%

Source: CNB and Author´s Computations

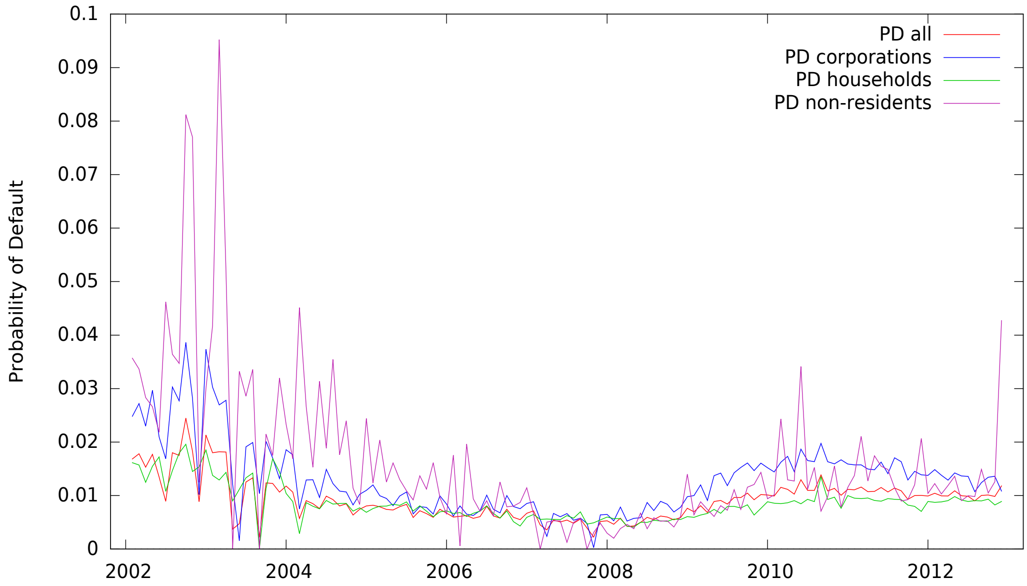
Figure 4.1: Evolution of Approximated PD Ratio of All Sectors during 01/2002–06/2013



Source: Author's computations.

All of the examined sectors (Figure 4.2) exhibited very similar behavior – higher volatility in the beginning of the observed period, a decrease till September 2008 followed by an increase up to the end of the observed period. Furthermore, none of PDs reached its pre-crisis maximum. First, the PD of non-residents seems to be the most volatile one with standard deviation 1.75% and it also has the largest average of 1.53% (Table 4.2), therefore the sector of non-residents can be consider to be the most risky one.

Figure 4.2: Evolution of Approximated PD Ratio of Sectors of Non-Financial Corporations, Households, Non-Residents, and PD of All Sectors Together during 2002–2013



Source: Author's Computations.

Table 4.2: Descriptive Statistics of Approximated PD ratios during 01/2002-06/2013

Variable	Mean	Min.	Max.	Std.Dev.
PD All	0.0093	0.0022	0.0245	0.0038
PD Corporations	0.0131	0.0003	0.0386	0.0066
PD Households	0.0085	0.0000	0.0196	0.0035
PD Non-Residents	0.0153	0.0000	0.0952	0.0175

Source: CNB and Author´s Computations

Second, in comparison to PD of Czech households and non-financial corporations, the PD of households has been reaching lower values than the PD of corporations as well as the volatility of PD of households is lower than the volatility of PD of corporations. Thus, the sector of households might be considered safer. On the other hand, the values of PD of households and PD of corporations seem to converge to the value of 1% in the end of observed period when the PD of non-residents went sharply up.

4.2 Macroeconomic Variables

The macroeconomic variables used are GDP growth, inflation, real exchange rate, interest rate, unemployment and indebtness of economy. The selection was based on data available and related papers (for example Virolainen (2004), Jakubík (2007), Simons & Rolwes (2009), or Jiménez & Mencía (2009)). Moreover, the relationship of GDP growth and PD and possible non-linearity are examined.

4.2.1 Gross Domestic Product (GDP) growth

One of the most important variables to determine PD is GDP. There are several papers which studies relationship between probability of default and business cycles. For example, Koopman & Lucas (2003) or Simons & Rolwes (2009) have found a negative relationship between PD and GDP growth. In the time of crisis when GDP growth goes down, the PD might be relatively high and vice versa. In the case of Czech Republic, Figure 4.3 shows possible negative relationship as well, but the slump of GDP growth in 2008 and 2012 is much larger and not proportional to the steady growth of PD. Moreover, the PD time series is more volatile in the beginning of the sample period when the GDP growth series is quite smooth.

GDP growth variable is extracted from time series of monthly real GDP data. Because the ARAD database offers quarterly data only, an interpolation was done – the method chosen for interpolation is Cubic Spline Interpolation using polynomial of 3rd degree referred for example in McKinley & Levine (1998).
The GDP growth at time t is defined as:

$$GDP\_gr_t = \frac{GDP_t - GDP_{t-1}}{GDP_{t-1}} \tag{4.2}$$

To test non-linearities in PD and GDP growth, Gasha & Morales (2004) identify threshold effect in NPL and GDP, i.e. they are looking for non-linearity in a data which causes different behaviour of NPL after reaching some value.

Figure 4.3: Comparison of Evolution of Approximated PD of All Sectors and GDP Growth during 01/2002–06/2013

Source: Author's Computations.

Because they use quarterly data and because of the definition of PD ratios, the adjusted SETAR model purposed by Gasha & Morales (2004) applied to monthly data is following:

$$PD_t = a + bGDP\_gr_t + c_1PD_{t-1} + c_2PD_{t-2} + c_3PD_{t-3} +$$
$$+ d_0T + d_1PD_{t-1}T + d_2PD_{t-2}T + d_3PD_{t-3}T + e_t$$

where PD_t is PD of whole economy or given sector at time t, GDP_gr_t is growth of GDP at time t, T is threshold, e_t is error term at time t and $a, b, c_1, c_2, c_3, d_0, d_1, d_2, d_3$ are parameters to be estimated.

The appropriateness of the model was tested by non-linearity tests which showed that the linear model might be better in case of PD of whole economy as well as in case of PD of given sectors. Hence, it is concluded that there should not be any non-linear reaction of PD to changes in its past values and values of GDP growth, and there should not be any threshold which divides the reaction of the approximated PD. This gives a reason to use linear models in the analysis.

4.2.2 Interest Rate

Another important determinant in each economy is interest rate. Generally, the higher real interest rate is, the more problems could the client have with repaying his debt. Hence, higher interest rate should influence the probability of default in the same way. If it is considered that low interest rates are usually connected with recession period and if it is assumed that PD is negatively correlated with GDP growth, the relationship between interest rates and PD might be negative as well. A negative relationship is also indicated by the results of Jakubík (2006) or Virolainen (2004). In the analysis, the interest rates are represented by monthly average of 1 month PRIBOR.

4.2.3 Exchange Rate

Exchange rate is one of key indicators for a small open economy as the Czech Republic. Furthermore, it has direct impact on loans in foreign currency. Since the Czech Republic does not have large ratio of loans determined in foreign currency on total loans (Czech National Bank 2013), an impact of changes in exchange rate should not have large impact on probability that the debtor will not pay back in case of unfavourable movement of exchange rate. Another result of exchange rate movement is change in the price of imports and exports. Czech economy is export-oriented (Tchaidze & Westin 2010), thus the depreciation of Koruna may increase the profit of Czech firms and the probability of default of these firms may become lower. On the other hand, the depreciation of Koruna might make the imports more expensive and firms can possibly have higher costs. Hence, the effect of the exchange rate on PD is not clear.

Simons & Rolwes (2009) did not find any significant influence of exchange rate on the PD, but using Czech data, Jakubík & Schmieder (2008) found an effect on exchange rates positive and significant. The exchange rate is defined as Koruna real exchange rate based on CPI where year 2010 is 100% and weights are based on foreign trade turnover.

4.2.4 Inflation

Changes in price level might be other important determinant of PD. Generally, inflation transfers wealth from lenders to borrowers as it is devaluating the loan during the time. Thus, there is an intuition for negative correlation between inflation and PD; the higher inflation should help borrowers to repay their

debts. This negative relationship was also found by Jakubík (2007).
Finally, the inflation represented by PPI index is defined as:

$$infl_t = \frac{PPI_t - PPI_{t-1}}{PPI_{t-1}} \tag{4.3}$$

4.2.5 Unemployment

This determinant should be significant especially for the sector of households. When the unemployment rate goes up, there might be larger amount of households which are not able to pay due to their reduced wealth. Thus, the positive sign between unemployment and PD is expected. Furthermore, data of unemployment rate was available since January 2005; therefore the analysis using unemployment has restricted number of observations.

4.2.6 Loans/GDP Ratio

The last examined determinant is the share of total loans of GDP or the indebtness of the economy. More loans in the economy might results in more defaults as the quality of borrowers will go down. On the other hand, there might be also a negative relationship between PD and indebtness of economy in case the number of loans is under its optimum level. The positive relationship was concluded by Jakubík & Schmieder (2008), however, Virolainen (2004) has found significant negative relationship between indebtness ratio and PD.
The variable is constructed using monthly data of total number of loans in the economy/sector and approximation of monthly time series of GDP which is interpolated from quarterly data using Cubic interpolation method.

Table 4.3: Descriptive Statistics of Employed Macroeconomic Variables during 01/2002–06/2013

	Mean	Standard Deviation	Minimum	Maximum
GDP growth	0.0021	0.0036	-0.0124	0.0090
PRIBOR 1M	2.0626	1.0687	0.3000	4.6600
Real Exchange Rate	91.6883	9.0305	76.9200	108.6000
Inflation - PPI Index	100.1914	0.4992	99.2000	103.0000
Unemployment	6.1644	1.0932	3.8000	8.1000
Loan/GDP	1.8242	0.4929	1.1699	2.5777

Source: CNB and Author's Computations.

4.3 Relationships between Macroeconomic Variables

The macroeconomic variables are often correlated which is not desirable as the assumption of independence of explanatory variables is an important part of plenty of models. Table 4.4 shows how the employed macroeconomic variables are correlated. This relationships hold not only for non-lagged variables, but also for lagged macroeconomic variables – all the lagged variables used in the further analysis perform similar results.

The correlation of Loans to GDP ratio with other variables might cause problems in estimations. The negative correlation of Loans to GDP ratio with interest rate and GDP growth is straightforward. Almost perfect correlation of Loans to GDP with real exchange rate could be explained as following. Higher real exchange rate might results in higher demand for foreign goods and services and lower demand for domestic goods and services. Thus, ceteris paribus, the net export goes down as well as the GDP (the denominator of the ratio) which results in higher Loans to GDP ratio. Therefore, this relationship might also explain some correlation between GDP growth and real exchange rate.

Finally, the last important correlation of variables is among the interest rate and unemployment. The negative correlation between them might be high as monetary policy authority usually lowers interest rates during the crisis when macroeconomic conditions are worsened and unemployment high and vice versa – monetary policy may try to increase interest rates higher in time of economic boom when there might start inflationary pressures leading to inflation above the target and when unemployment rate is usually lower.

Table 4.4: Correlation of Employed Macroeconomic Variables during 01/2002–06/2013

	Inflation	PRIBOR1M	Real Ex. Rate	Unempl.	GDP Gr.	Loans/GDP
Inflation	1					
PRIBOR 1M	-0.0263	1				
Real Ex. Rate	-0.0486	-0.1944	1			
Unemployment	0.0438	-0.8606	-0.072	1		
GDP Growth	0.0166	0.1342	-0.5042	0.0894	1	
Loans to GDP	-0.0165	-0.4465	0.8822	0.137	-0.677	1

Source: CNB and Author's Computations.

Chapter 5

Results of Empirical Analysis

First, the ARIMA analysis of dependence of the PD on its own values is tested. To analyze an effect of macroeconomic variables on the PD of whole economy as well as on the PDs of chosen sectors, the latent factor model and FAVAR model described in detail above were employed. The model selection was mainly influenced by the availability of data.

5.1 ARIMA Analysis

To reveal a dependence of the PD ratio on its past values, the basic ARIMA modelling was done. Information Criteria recommend ARIMA (3, 0, 0) model, but the residuals of estimated model do not satisfy the assumption of its normality (test stat 489.5, p-value 0.000), moreover, ARCH-LM test might result in rejecting of the null hypothesis of no ARCH effects at 1% level of confidence (test stat 19.85, p-value 0.0005). These violations of assumptions of the ARIMA models might not influence significance of variables, but the coefficients might be biased, hence the further analysis by more advanced methods and using more variables should be done. Results (Table 5.1) show a positive significance of the lagged values of PD ratio of whole economy of one month (t-stat 5.83) and three months (t-stat 4.76) lagged; hence the PD might be dependent on its past values up to the one quarter delay.

A dependence on past values was also tested for given sectors. For the PD of sector of corporations, it was also recommended to employ 3 lags of dependent variable by the information criteria. The residuals from ARIMA (3, 0, 0) estimation do not satisfy the assumption of no ARCH effect (test stat

Table 5.1: Approximated PD of All Sectors - Results of ARIMA (3, 0, 0) Estimation

	Coefficients	Std. Errors	T-Ratio	Approx. Prob.
AR1	0.47315608	0.08113012	5.83206	0.00000
AR2	0.01618782	0.09146434	0.17699	0.85980
AR3	0.38663125	0.08128613	4.75642	0.00001
CONST	0.00998365	0.00161004	6.20086	0.00000

Source: Author's Computations.

Table 5.2: Approximated PD ratio of non-financial corporations - Results of ARIMA (3, 0, 0) Estimation

	Coefficients	Std. Errors	T-Ratio	Approx. Prob.
AR1	0.42367148	0.07851668	5.39594	0.00000
AR2	0.03044743	0.08704127	0.34980	0.72706
AR3	0.44789980	0.07858274	5.69972	0.00000
CONST	0.01426879	0.00315740	4.51916	0.00001

Source: Author's Computations.

23.26, p-value 0.0001) and the null hypothesis of normality (test stat 393.5, p-value 0.0000) was rejected as well, both at 1% level of confidence. On the other hand, autocorrelation does not seem to be violated and results (Table 5.2) show significantly positive dependence on one month (t-stat 5.4) and three months (t-stat 5.7) lagged values. The violation of assumption should not change significance of strongly significant variables, but the estimated coefficient might be biased. The approximated PD of corporations exhibits very similar results as the ARIMA case of the PD of all sectors when the two months lagged values (t-stat 0.35 in comparison to the t-stat 0.18 of all sectors modelled together) are not taken into account when the most recent values and tree months lagged values seem to be important determinant of the PDs.

The PD of households was modelled by ARIMA (4, 0, 0) according to the information criteria. The autocorrelation check and no ARCH effect test (test stat 5.6, p-value 0.23) performed well on 1% confidence level, but the normality was rejected (test stat 4998.8, p-value 0.000). Hence, the coefficients might be biased. Compared to the non-financial corporations, the most important determinant is also one month lagged variable (t-stat 4.79), but the approximated PD of households might be more persistent as the results shows significance of four month lagged dependent variable (t-stat 5.2) which is not significant or modelled for any other sector.

Table 5.3: Approximated PD of Households - Results of ARIMA (4, 0, 0) Estimation

	Coefficients	Std. Errors	T-Ratio	Approx. Prob.
AR1	0.40831218	0.08525428	4.78934	0.00000
AR2	0.01330775	0.08971310	0.14834	0.88231
AR3	0.23736465	0.09042446	2.62500	0.00974
AR4	0.26102656	0.08537173	3.05753	0.00273
CONST	0.00959409	0.00184672	5.19520	0.00000

Source: Author's Computations.

Table 5.4: Approximated PD of Non-Residents - Results of ARIMA (3, 0, 0) Estimation

	Coefficients	Std. Errors	T-Ratio	Approx. Prob.
AR1	0.46133236	0.08611547	5.35714	0.00000
AR2	-0.02994378	0.09591192	-0.31220	0.75540
AR3	0.29889040	0.08632850	3.46224	0.00073
CONST	0.01711763	0.00373139	4.58747	0.00001

Source: Author's Computations.

Finally, ARIMA (3, 0, 0) was employed in case of non-residents. Except of rejection normality (test stat 2964.3, p-value 0.000) of residual and possible presence of ARCH effect (test stat 12.08, p-value 0.02) on 5% confidence level, there might be also a problem with autocorrelation which may negatively influence a significance of lagged variables. But one month lagged (t-stat 5.36) and three month lagged variable (t- stat 3.46 shows a smallest dependence on that values among sectors) seem to be quite significant. Hence, the violation should not affect the conclusion that there might be a significant dependence on past values up to the one quarter, except of the two month lagged values as in the case of sector of non-financial corporations and all sectors modelled together. On the other hand, the estimated coefficients might be biased. Therefore, the further analysis of the PD should be done.

To conclude, all of the approximated PDs seem to be very dependent on its past values, especially, on the one month and three months lagged ones. The largest persistence (up to 4 months) might have sector of households when the other two sectors perform dependence up to third month, but any of the sectors does not depend on the second month lagged vales. Finally, some assumptions might be violated, but the significance of lagged values should be robust to them, although the estimated coefficients might be biased, hence those models might not be appropriate for PD modelling and more variables than past values only should be employed.

5.2 One Latent Factor Model

One latent factor model was estimated by maximum likelihood estimator. According to Green (2012), there is an assumption of independent and identically distributed variables. Therefore, the variables of real exchange rate (positively correlated GDP growth), indebtness of the economy (negatively correlated with GDP growth as the loans in the economy are stable and the indebtness ratio react primary on changes in GDP growth) and interest rate (negatively correlated with unemployment) had to be excluded from all estimated models.
The choosing of number of lags in each sector/model was based on appropriateness and model fit. This was done by minimizing the information criteria – Akaike and Bayesian information criteria.
All of the models performed very similar results in case the unrestricted dataset was employed. Hence, there should not be any significant differences in results implied by those two datasets. Thus, the results are robust to those changes.
All of the models were also tested by Wald test. The Wald test is a linear test with the null hypothesis that all the coefficients are equal to zero. The null hypothesis was rejected at 1% level of significance for all of the models (p-values 0.000). Therefore the models should not be misspecified.

5.2.1 PD of All Sectors

For the PD of all sectors, the chosen model is following:

$$pdall_t = F_N(\alpha + \beta_1 infl_{t-3} + \beta_2 GDPgr_{t-3} + \beta_3 unempl_t + \beta_4 dummy + \beta_5 \rho) \tag{5.1}$$

where $pdall_t$ is PD of all sectors at time t, F_N is function of cumulative standard normal distribution, $infl_t$ represents inflation of Czech economy at time

t, $GDPgr_t$ is GDP growth at time t, $unempl_t$ unemployment at time t, $dummy$ is dummy variable which equals to 1 when time t reaches September 2008 and 0 otherwise, and ρ is latent factor with standard normal distribution.
Maximum likelihood estimation method was employed to estimate the model. The results are illustrated in Table 5.5. It is indicated that three months lagged inflation is not significant determinant of probability of default (z-stat -1.35) at 10% level of significance. On the other hand, the rest of the macroeconomic variables (three months lagged GDP growth and unemployment) seem to be strongly significant if the 1% level of significance is considered. Unemployment is suggested to be even more important in the PD of all sectors modelling (z-stat 14.45) than three months lagged GDP growth (z-stat -4.19). Moreover, the significance of dummy variable (z-stat 4.96) indicates that the financial crisis might influence the Czech PD of all sectors in September 2008. Finally, the latent variable is not proven to be significant determinant of the PD (z-stat 0.01).

Table 5.5: Results of Latent Factor Model Estimation - PD of All Sectors

Variable (# lags)	Coefficient	Robust Std. Err	z	P-Value	[95% Conf.	Interval]
Constant	-2.872493	.0347935	-82.56	0.000	-2.940687	-2.804299
Inflation (-3)	-1.309913	.9715035	-1.35	0.178	-3.214024	.5941992
GDP growth (-3)	-5.635572	1.34471	-4.19	0.000	-8.271155	-2.99999
Unemployment	.0773301	.0053512	14.45	0.000	.0668419	.0878182
Dummy	.0754915	.0152127	4.96	0.000	.0456752	.1053078
Rho	.000051	.0059176	0.01	0.993	-.0115473	.0116493

Source: Author's Computations.

Because of the normal distribution used in modelling the PD, the marginal effects have to be computed to reveal how the PD reacts on moves in the explanatory variables on average. The marginal effects are displayed in the Table 5.6. Positive relationship is indicated by the analysis between the PD and unemployment. Furthermore, a negative relationship between the PD and one quarter lagged GDP growth was also suggested by the analysis. Due to the strong correlation of excluded variables with included significant variables, there might also be positive relationships between the PD and real exchange rate and the PD and indebtness. Negative relationship between the PD and interest rate might also exist as there is a correlation between interest rate and unemployment. Lastly, because the coefficient of dummy variable is positive, the financial crisis might influence the PD of all sectors positively overall.

Table 5.6: Marginal Effects - Latent Factor Model Estimation - PD of All Sectors

Variable (# lags)	dy/dx	Delta Method Std. Err	z	P-Value	[95% Conf. Interval]	
Inflation (-3)	-.0293849	.0218265	-1.35	0.178	-.072164	.0133943
GDP growth (-3)	-.1264211	.0296396	-4.27	0.000	-.1845137	-.0683286
Unemployment	.0017347	.000123	10.14	0.000	.0014936	.0019759
Dummy	.0016935	.0003282	5.16	0.000	.0010502	.0023367

Source: Author's Computations.

Finally, the forecasts were made six periods ahead and the performance of the model was assessed. The model seems to overestimate true values of the PD of all sectors, but according to the forecasting measures, the model might still be appropriate to model the PD of all sectors. The mean forecast error is almost -0.0016 and the RMSE almost 0.0016 (actual values of the PD are fluctuating around 0.01) when the value of mean absolute percentage error reaches almost 15.03%.

5.2.2 PD of Non-Financial Corporations

For the PD of non-financial corporations, the chosen model is following:

$$pdcorp_t = F_N(\alpha + \beta_1 infl_{t-3} + \beta_2 GDPgr_t + \beta_3 unempl_t + \beta_4 dummy + \beta_5 \rho) \tag{5.2}$$

where $pdall_t$ is PD of all sectors at time t, F_N is function of cumulative standard normal distribution, $infl_t$ represents inflation of Czech economy at time t, $GDPgr_t$ is GDP growth at time t, $unempl_t$ unemployment at time t, $dummy$ is dummy variable which equals to 1 when time t reaches September 2008 and 0 otherwise, and ρ is latent factor with standard normal distribution.
Results of the model estimation are given in the Table 5.7. The results are quite similar to the results of the estimation of the PD of all sectors. All of the variables except of three months lagged inflation (z-stat -0.97) seem to be significant determinants of the PD of non-financial corporations at the significance level of 1%. In comparison, unemployment (z-stat 9.14) is also indicated to be more significant than GDP growth (z-stat -3.34) when the PD of non-financial

corporations is modelled. Furthermore, analysis indicated that the sector of non-financial corporations should be influence by the crisis as the dummy variable is quite significant (z-stat 4.53) as well. Finally, the latent variable is not significant (z-stat -0.23) when analyzing sector of non-financial corporations.

Table 5.7: Results of Latent Factor Model Estimation - PD of Non-financial Corporations

Variable (# lags)	Coefficient	Robust Std. Err	z	P-Value	[95% Conf. Interval]	
Constant	-2.844251	.0565806	-50.27	0.000	-2.955147	-2.733355
Inflation (-3)	-1.326515	1.360852	-0.97	0.330	-3.993736	1.340705
GDP growth	-6.681971	2.002634	-3.34	0.001	-10.60706	-2.756882
Unemployment	.0907329	.0089949	9.14	0.000	.0731033	.1083625
Dummy	.1031085	.0227379	4.53	0.000	.058543	.147674
Rho	-.0020561	.0091136	-0.23	0.822	-.0199184	.0158062

Source: Author's Computations.

Marginal effects (Table 5.8) indicate negative relationship between GDP growth and the PD and positive relationships between the PD and unemployment and between the PD and dummy variable. Hence, the analysis suggests that the crisis as well as an increase in unemployment might impact the value of the PD of non-financial corporations positively when an effect of an increase in GDP growth is negative and vice versa. Finally, due to the correlation between variables, the PD might be also positively influenced by real exchange rate. Moreover, the PD might also negatively react on positive changes in interest rate and in indebtness.

Table 5.8: Marginal Effects - Latent Factor Model Estimation - PD of Non-Financial Corporations

Variable (# lags)	dy/dx	Delta Method Std. Err	z	P-Value	[95% Conf. Interval]	
Inflation (-3)	-.038608	.0397061	-0.97	0.331	-.1164306	.0392146
GDP growth	-.1944777	.058103	-3.35	0.001	-.3083574	-.0805979
Unemployment	.0026408	.00027	9.78	0.000	.0021116	.0031699
Dummy	.003001	.0006384	4.70	0.000	.0017498	.0042521

Source: Author's Computations.

Forecast six periods ahead seems to overestimate the true values of the PD of the non-financial corporations. This was indicated by forecasting measures which also assess appropriateness of the model. The RMSE is almost 0.0037, mean forecast error over -0.0029 (actual values are oscillating around 0.14), and mean absolute percentage error almost 24.14%. Thus, the model might be applicable, although it might not be the best model of the PD of non-financial corporations.

5.2.3 PD of Households

For the PD of households, the chosen model is following:

$$pdhh_t = F_N(\alpha + \beta_1 infl_{t-6} + \beta_2 GDPgr_{t-3} + \beta_3 unempl_t + \beta_4 dummy + \beta_5 \rho) \tag{5.3}$$

where $pdall_t$ is PD of all sectors at time t, F_N is function of cumulative standard normal distribution, $infl_t$ represents inflation of Czech economy at time t, $GDPgr_t$ is GDP growth at time t, $unempl_t$ unemployment at time t, $dummy$ is dummy variable which equals to 1 when time t reaches September 2008 and 0 otherwise, and ρ is latent factor with standard normal distribution.
Results (Table 5.9) of application of maximum likelihood estimation on latent factor model for households indicates that all the macroeconomic variables except three months lagged GDP growth (z-stat -1.48) should be significant on 10% level of significance. The strongest significance is indicated for unemployment (z-stat 14.2) as six months lagged inflation is not anymore significant determinant of the PD of households when 5% level of significances is assumed. Therefore, unemployment should be the most important determinant of the PD of households according to the latent factor model. Furthermore, for the sector of households, dummy variable is not significant (z-stat -0.41). Hence, the analysis indicates that the financial crisis should not significantly influence the PD of households. Finally, the latent factor does not seem to be significant determinant of the PD (z-stat 0.56).

Marginal effects of the latent factor model for the sector of households (Table 5.10) indicate positive relationships between the PD and six months lagged inflation and the PD and unemployment. Therefore, the households might be less able to repay their debts in conditions of higher unemployment and higher inflation. As a strong correlation between unemployment and interest rate was found, there might also be a negative relationship between the PD of house-

Table 5.9: Results of Latent Factor Model Estimation - PD of Households

Variable (# lags)	Coefficient	Robust Std. Err	z	P-Value	[95% Conf. Interval]	
Constant	-2.776775	.0250125	-111.02	0.000	-2.825799	-2.727751
Inflation (-6)	1.109181	.6689047	1.66	0.097	-.201848	2.42021
GDP growth (-3)	-1.557832	1.050398	-1.48	0.138	-3.616574	.5009103
Unemployment	.0557551	.0039265	14.20	0.000	.0480592	.063451
Dummy	-.0044189	.0108132	-0.41	0.683	-.0256123	.0167745
Rho	.0025257	.0043774	0.58	0.564	-.0060538	.0111053

Source: Author's Computations.

holds and interest rate.

Table 5.10: Marginal Effects - Latent Factor Model Estimation - PD of Households

Variable (# lags)	dy/dx	Delta Method Std. Err	z	P-Value	[95% Conf. Interval]	
Inflation (-6)	.0228948	.0138761	1.65	0.099	-.0043018	.0500914
GDP growth (-3)	-.0321555	.0215835	-1.49	0.136	-.0744583	.0101473
Unemployment	.0011509	.0000844	13.64	0.000	.0009855	.0013162
Dummy	-.0000912	.0002237	-0.41	0.684	-.0005297	.0003473

Source: Author's Computations.

Forecast error measurement shows that the model overestimates the true values of the PD of households. But according to the forecasting measure, the model seems to be able to capture evolution of the PD quite well. The RMSE is almost 0.0011, mean forecast error less than -0.0009 (to comparison, actual values are around 0.009) and mean absolute percentage error over 10.52%.

5.2.4 PD of Non-Residents

For the PD of non-residents, the chosen model is following:

$$pdnon_t = F_N(\alpha + \beta_1 infl_{t-6} + \beta_2 GDPgr_{t-3} + \beta_3 unempl_{t-3} + \beta_4 dummy + \beta_5 \rho) \tag{5.4}$$

where $pdall_t$ is PD of all sectors at time t, F_N is function of cumulative standard normal distribution, $infl_t$ represents inflation of Czech economy at time t, $GDPgr_t$ is GDP growth at time t, $unempl_t$ unemployment at time t, $dummy$

is dummy variable which equals to 1 when time t reaches September 2008 and 0 otherwise, and ρ is latent factor with standard normal distribution.
The results of the model of the PD of non-residents are summarized in Table 5.11. Three month lagged GDP growth (z-stat -0.63) is not indicated to be significant determinant of the PD of non-residents. The only strongly significant variable and probably the most important one is three month lagged unemployment (z-stat 6.73). On the other hand, a significance of unemployment is still lower than significance of unemployment (as the most significant determinants of other sectors) when the PDs of other sectors are modelled. At 10% level of significance, the variables six months lagged inflation (z-stat -1.95) and dummy variable (z-stat 1.69) might be significant determinants of the PD of non-residents as well. Thus, it confirms the hypothesis that the financial crisis might also have an influence on the sector of non-residents. Finally, the latent factor (z-stat -0.63) should not be significant determinant of the examined PD as well.

Table 5.11: Results of Latent Factor Model Estimation - PD of Non-Residents

Variable (# lags)	Coefficient	Robust Std. Err	z	P/Value	[95% Conf. Interval]	
Constant	-3.156654	.1313353	-24.04	0.000	-3.414066	-2.899241
Inflation (-6)	-4.503094	2.309025	-1.95	0.051	-9.028699	.0225115
GDP growth (-3)	-2.410626	3.797557	-0.63	0.526	-9.8537	5.032449
Unemployment (-3)	.1367235	.0203117	6.73	0.000	.0969134	.1765337
Dummy	.1003985	.0595181	1.69	0.092	-.0162549	.2170518
Rho	-.0101061	.0159705	-0.63	0.527	-.0414076	.0211954

Source: Author's Computations.

Marginal effects (Table 5.12) indicate a negative relationship between the PD and six months lagged inflation and positive ones between the PD and dummy variable and the PD and unemployment. Also, the relationship between dummy variable and the PD might indicate that the PD of the sector of non-residents was also positively influenced by the financial crisis in September 2008. Lastly, the correlation between unemployment and interest rate might indicate a negative impact of interest rate on the PD of non-residents.

Forecast measurement also indicates that the latent factor model tends to overestimate the PD of non-residents. The forecast measures do not perform

Table 5.12: Marginal Effects - Latent Factor Model Estimation - PD of Non-Residents

Variable (# lags)	dy/dx	Delta Method Std. Err	z	P-Value	[95% Conf. Interval]	
Inflation (-6)	-.1257988	.0667336	-1.89	0.059	-.2565942	.0049966
GDP growth (-3)	-.0673434	.1063947	-0.63	0.527	-.2758733	.1411864
Unemployment (-3)	.0038195	.0006016	6.35	0.000	.0026403	.0049987
Dummy	.0028047	.0016522	1.70	0.090	-.0004335	.006043

Source: Author's Computations.

the best values between all the examined sectors, but it might be still quite good model for the PD of nonresidents as the value of RMSE is over 0.0051, mean forecast error is over -0.0007 (actual values are fluctuating around 0.016). On the other hand, mean absolute percentage error is almost 32.12%.

5.2.5 Conclusions

Results of all of the models seem to confirm the importance of macroeconomic variables in the PD modeling. On the other hand, the significance of latent variable was not proved in any sector. Therefore there might be no remaining effects which could be explained by unobserved variable if the normality of the variable is assumed.

The general macroeconomic index, the GDP growth of the Czech Republic, seems to be significant variable in determining the PD except the sector of non-residents. The PD of non-residents seems to be independent on that measure which is reasonable as economic activity of non-residents might not be related to Czech GDP. Generally, the results performed a negative sign of coefficient of the GDP which was initially expected. Thus, the analysis suggests that the lower is growth in GDP, the higher is the PD and vice versa.

Inflation was significant only when the PD of households and the PD of non-residents were modelled and the level of significance was assumed to be 10%. The signs are, though, different as the PD of households seems be influenced positively and PD of non-residents negatively. Hence, higher inflation might help non-residents to repay their debts as the real value of the debt might go down. On the other hand, higher households´ default rates are indicated to be connected with higher inflation.

By the analysis, the most significant macroeconomic determinant of all of the PDs is indicated to be unemployment with expected positive sign of its coefficient. Therefore, the analysis indicates that the higher unemployment should increase the PD of whatever sector and vice versa. Even the PD of non-residents seems to be strongly connected to Czech unemployment rate.

Dummy variable, the crisis indicator, is significant when all the sectors together, non-residents and non-financial corporations were analyzed. Except of the households sector, all the sectors seems to be negatively hit by the crisis as the sign of the coefficients of dummy variables was found to be positive. Thus, the sector of households is the only sector which is suggested to not react on the circumstances of September 2008.

Finally, all of the models performed quite well in forecasting and fitting future values (Table 5.13). The best fit and the lowest forecast errors are performed by the sector of households followed by the PD of all sectors modelled together. Thus, the latent factor model should explain the movements in those PDs quite well. On the other hand, for the sector of non-residents, there might also be a better model than the one using Czech macroeconomic variables only as the MAPE is almost 32.12%.

Table 5.13: Forecast Measures Applied on Six Periods ahead Forecasts of Latent Factor Model Divided by Sectors

	MAPE	MFE	RMSE
PD of All Sectors	15.0258%	-0.001575	0.001595
PD of Corporations	24.1384%	-0.002942	0.003693
PD of Households	10.5230%	-0.000882	0.001064
PD of Non-Residents	32.1198%	-0.000773	0.005102

Source: Author's Computations.

5.3 FAVAR

First, the optimum number of principal components was determined at 3 as those components are explaining almost 90% of all variance when employing 2 components would explain only 70% of variance o macroeconomic variables which might not be enough for the following analysis. The component loadings and share on variance explanation of the components is shown in Table 5.14 and 5.15.

Table 5.14: Principal Components Analysis - Properties of Components

Component	Eigenvalue	Proportion	Cumulative
1	2.52	0.4225	0.4225
2	1.31	0.3022	0.7247
3	0.9954	0.1659	0.8906
4	0.5166	0.0861	0.9767
5	0.0992	0.0165	0.9932
6	0.0406	0.0068	1.0000

Source: Author's Computations.

Table 5.15: Principal Components Analysis - Composition of Components

	PC1	PC2	PC3	PC4	PC5	PC6
Inflation	-0.016	0.080	-0.996	0.041	-0.000	0.014
Indebtness	0.605	-0.128	-0.024	0.158	0.269	-0.721
GDP growth	-0.446	0.287	0.061	0.810	0.139	-0.198
Real Exchange Rate	0.519	-0.273	-0.000	0.558	-0.336	0.481
Interest Rate	-0.351	-0.596	-0.047	0.024	-0.596	-0.405
Unemployment	0.204	0.682	0.046	-0.071	-0.663	-0.214

Source: Author's Computations.

Second, the number of lags used when performing VAR was also selected by minimizing the following information criteria: Akaike Information Criterion, Hannan-Quinn Criterion and Schwarz Criterion. For monthly data, the most relevant should be Akaike Criterion as recommended for example by Ivanov & Kilian (2005). In all of the cases (for all sectors as well as for the aggregated PD), the maximum number of lags recommended by all of the criteria was 10 and this number is also employed in the analysis. Hence, it is not expected that the dependency in variables should exceed one year.

5.3.1 PD of All Sectors

First, after performing Factor Augmented Vector Autoregression, tests of assumptions to confirm validity of the results were done. The null hypothesis of autocorrelation was rejected on 10% confidence level (p-values 0.954, 0.973, 0.74, and 0.659), as well as the hypothesis of remaining ARCH effect (p-values 0.13, 0.665, 0.51, and 0.6). Moreover, normality of residuals is also not rejected on 5% confidence level (p-value 0.0606). Finally, the stability and stationarity condition of VAR models has to hold. That means that the inverse roots of characteristic polynomial have to be inside a unit circle (Lütkepohl 2005). Because, this condition is also satisfied, the VAR system should not perform any spurious regressions and forecasts.

Variance decomposition shows that the moving in PD is mostly influenced by its own shocks. This influence has declining trend over time (but still more than 50% of variance is explained by itself) and shocks in principal components seem to have much more influence than in the time when shock appeared. Especially, a shock in first principal component connected with general macroeconomic conditions (correlated mainly with GDP growth, real exchange rate, and indebtness) might be important in determining movements in the PD. A shock in "nominal component" (correlated with inflation) could explain about 15% of movement in the PD when shocks in the second component (mainly correlated with unemployment and inflation) seem to have not very significant influence.

Another tool often employed to interpret VAR methodology is impulse responses analysis. The impulse responses analysis simulates shock in variable and track an impact on other variables. First, one standard deviation positive shock in the PD was simulated. According to the analysis, this shock might cause an increase in the PD followed by a decreasing oscillation to zero. Thus in long term, this shock should not have a significant effect.

Second, the analysis suggests that the same shock in first component, i.e. positive shock in indebtness and real interest rate and negative one in GDP growth, should result in the decline in PD first, but then, it seems to have positive impact on PD followed by a convergence to zero in long term. Thus, there might be positive relationship between the PD and GDP growth and negative one between the PD and interest rate and indebtness in medium term period and the opposite ones in short term period when in long term, all the relationships might disappear.

Third, one standard deviation positive shock in second component causes upward movement first, afterwards, the value of the PD seems to converge to a level below the original value and then converge back to zero in long term. Hence, there might be negative relationship between PD and unemployment and a positive one between the PD and interest rates in medium term and in the short term, the opposite ones.

Fourth, analysis indicates that the third principal component might have nega-

tive effect on the PD after initial increase in value of the PD. Specifically – due to almost perfect negative correlation with inflation – higher inflation should result in increasing the PD over time and it should persist in long term as well. Forecast for following six periods was made and compared to the actual data. The FAVAR seems to overestimate actual data. Thus, the real overall PD might be lower than forecasted. Mean absolute percentage error is around 13% when the root square mean error (RMSE) is around 0.0016 and mean forecast error equals to the value of -0.0013 (when the level of actual PD is around 0.01). Therefore, the model might quite well capture the actual values.
Finally, the robustness of the results was checked. The dataset containing outliers gives the same results as the dataset without outliers; therefore, the restrictions in data do not influence an accuracy of results.

5.3.2 PD of Non-Financial Corporations

First, the tests to confirm reliability of the model were done. The null hypothesis of no autocorrelation (p-values 0.817, 0.559, 0.627, and 0.599) is not rejected on 10% significance level as well as the null hypotheses of no remaining ARCH effect (p-values 0.81, 0.15, 0.85, and 0.49) and normality (p-value 0.4894). Moreover, the stability test has not shown any inverse root of characteristic polynomial outside a unit circle. Thus, the assumptions should be satisfied and the results of the model should be reliable.
Variance decomposition of the PD of non-financial corporations indicates large dependence of movements of the PD on its past shocks as it explains more than 60% of those movements. In the group of principal components variables, the most important determinant seems to be the first principal components ("general economic conditions" component) when the influence of the second principal component (component of interest rate and unemployment) should be minimal. On the other hand, the importance of shocks in third component ("component of inflation") in explanation of movements in the PD third is rising over time and in the end of the examined period, its shocks explain more than the shocks in the first component.
Impulse responses of the PD in reaction to one standard deviation shock in the PD itself shows stable increase in the value of PD after initial fluctuations. In long term, the value starts to decrease again and cross the zero. Thus, the long term relationship between the PD and the third component might be negative

when the short and medium term ones appear to be positive.
According to the analysis, the PD should react to the one standard deviation positive shock in the first principal component by decreasing its value immediately followed by an increase and it should come back to its initial value after some fluctuations. Thus, shocks in indebtness, GDP growth, and real exchange rate of the Czech economy should influence the PD of non-financial corporations in both ways, but the influence might be only short or medium term.
Reaction of the PD to the shock in second component is opposite to the first component one. After the shock, the value of the PD should increase. This increase should be followed by decrease and further increase. Finally, the PD should return to its initial value. This indicates that unemployment and interest rate might not have any long term effect on the PD as well, but their shocks can temporarily move by the PD in both ways.
Finally, according to the impulse responses analysis, the shock in the third component seems to have negative impact on the PD. Thus, the decrease in inflation may cause constant decline in the PD of non-financial corporations. However, the PD seems to increase back to zero and increase above zero afterwards. Therefore, the effect of shock in inflation on the PD might be negative in long run.
As the forecast errors are positive, the model tends to underestimate true values of the PD of corporations. Thus, this model might also underestimate loses for creditors which would use it. Moreover, the errors are higher than errors got in the case of the model of PD of all sectors – RMSE and mean forecast error are around 0.004 (actual values are around 0.014) and mean absolute percentage error is more than 23%, but it could be still considered as a good fit of data.
Finally, the analysis with data containing outliers performed the same results as the analysis with the data without outliers. Thus, the analysis might not be influenced by the restriction in employed data.

5.3.3 PD of Households

First, the tests of the model assumptions need to be done. The null hypotheses of no autocorrelation (p-values 0.878, 0.746, 0.827, and 0.973) and no remaining ARCH effect (p-values 0.58, 0.145, 0.4, and 0.99) were not rejected at 10% level of significance. Furthermore, the inverse roots of characteristic polynomial were inside the unit circle, thus, stability of coefficients and stationarity was not rejected as well. On the other hand, normality was rejected (p-value 0.000) at the 1% level of significance. According to Luetkepohl (2011), the violation of normality should not have an impact on most of the procedures connected to VAR; it may be indication of some non-linearities. Following this idea, the non-linearities in the PD of household were previously tested and rejected. Hence, the problems of the model might be with confidential intervals only according to Luetkepohl (2011). To assure there are no more problems with estimation errors, the estimation using robust standard errors was employed.

The variance decomposition reflects larger importance of other components in explaining movements in the PD than the PD itself in comparison to the other sectors. The shocks in the PD are still the most important determinants in the PD movement in time, but it declines up to the 40% during the observed period when other components are explaining around 20% each. Thus, the macroeconomic variables are indicated to be important determinants of the PD movements. Impulse responses analysis suggests that the one standard deviation shock in the PD of households immediately increases of the PD, but afterwards, the value of the PD decreases and converges to zero. Therefore, the shock in the PD might not be very persistent and should not impact itself in long run significantly.

The shock of one standard deviation applied to the first principal component results in an initial growth replaced by a decrease few periods ahead, but in the end of the observed period, the value of the PD seems to converge to zero. Thus, the GDP growth, indebtness of the economy and real exchange rate do not seem to be significant determinants in the long run, although, in short and medium run, there might be a positive relationship between the PD and indebtness of the economy and the PD and real exchange rate and a negative one between the PD and GDP growth.

The opposite case in terms of relationships between variables to a shock in the first principal component is indicated by impulse responses analysis of the shock in the second component to the PD of households. An initial decline is followed by growth which ends by fluctuating around zero. Thus, the interest rate and unemployment might not have any significant influence on the PD of Czech households in long run as well, but in the short or medium periods, a negative relationship between the PD and unemployment and a positive one between the PD and interest rate might hold.

According to the impulse responses analysis, the principal component highly correlated with inflation seems to have similar effect as the one from analysis of non-financial corporations. The initial growth in the PD after one standard

deviation shock in the third principal component continues by declining trend in the PD of households. Therefore, there could be significant positive relationship between inflation and the PD of households in long run while a shock in inflation might initially influence the PD negatively.
Forecast errors shows overestimation of the PD of households by the FAVAR model. The ratios measuring the forecasting errors are slightly higher than the ratio resulting from the analysis of all sectors, but they are lower than ratios indicated by analysis of non-financial corporations. The mean absolute percentage error is over 18.5%, RMSE 0.0018, and mean forecasting error is almost -0.0016 (when actual values are oscillating around 0.009), thus the model might be still acceptable.
Finally, robustness was also checked - both datasets performs very same results; therefore the restrictions should not influence results also in case of households analysis.

5.3.4 PD of Non-Residents

All the tests except of the normality performed well. The null hypotheses of no autocorrelation (p-values 0.928, 0.955, 0.837, and 0.987) and no remaining ARCH effect (p-values 0.92, 0.27, 0.55, and 0.99) were not rejected on the 10% level of significance; all the inverse roots of characteristic polynomial were inside a unit circle. Hence, the stability and stationarity condition hold as well. Because of the violation of normality assumption (p-value 0.0001), the estimation with robust standard errors was employed. Moreover, due to the possibility of non-linear dependencies in the PD, the non-linearity in the PD was previously tested and successfully rejected, so the violation of the normality of residuals should have an impact on confidence intervals only as suggested by Luetkepohl (2011).
Variance decomposition indicates that the most of the movement in the PD of non-residents is influenced by the shocks in the PD itself. Regarding the macroeconomic variables, shocks in the first two principal components seem to have only marginal effect on those movements. On the other hand, the third principal component appears to have some explanation power. Thus, similarly to the previous cases, shocks in inflation should have an effect on the PD when shocks in the rest of the macroeconomic variables might not affect the movements in the PD of non-residents according to the analysis.

Impulse responses analysis suggests that the one standard deviation positive shock in the PD should increase its value immediately and converge to zero afterwards. Therefore, there should be no long term effect of the shocks in the PD itself.
Second, the shock in the first component seems to create a cycles which should slowly converge to zero as well, but it takes longer period of time than the convergence when shock is applied into the PD itself. Therefore, analysis suggests that the indebtness of the economy, GDP growth, and real exchange rate should not have a long-term impact on the PD of non-residents. However, the period during which they might cyclically affect the PD might be quite long.
Third, analysis indicates that the positive shock in the second component creates a decrease in the PD initially. Then, the value of the PD seems to rise back to zero with high frequency of cycles. Therefore, there might be a short term negative relationship between the PD and unemployment and a positive one between the PD and interest rate, but in long term, the relationships should disappear.
Further, initially positive impact of the shock in the third component results in decrease afterwards, so the effect on the PD seems to be finally negative according to the analysis. Hence, a growth in inflation might cause immediate decrease in the PD of non-residents in short term when in long term, it should create an increase of the value of the PD of non-residents.
Forecasting by the chosen model shows that the FAVAR model overestimates the values of the PD. The values of RMSE (over 0.0064) and mean forecasting error (almost -0.003) are not very large as the actual values of the PD are over 0.016, but the mean absolute percentage error is not small (almost 40%). Therefore, the model might not be the best choice in forecasting the PD of non-residents.
Finally, robustness in terms of restriction of the data was checked as well. The results of the robustness check suggest that there should not be any significant differences in the results between the two datasets.

5.3.5 Conclusions

Analysis indicates that the all of the PDs´ movements are strongly dependent on their own shocks. On the other hand the macroeconomic determinants are able to explain some of the movements as well. The sector most dependent on its own shocks should be the sector of non-residents. Hence, the dependency of this sector on macroeconomic environment of the Czech Republic might be very small as only inflation seems to have a significant influence on the PDs of all mentioned sectors. Nevertheless, sector of households performed relatively low dependence on its own shock and all the macroeconomic variables together were able to explain more than one half of its movements. Generally, all the sectors performed certain dependence on inflation, most of them (except of the sector non-residents) also performed some dependence on GDP growth, indebtness of the economy, and real exchange rate, but unemployment and interest rate do not seem to be able to explain movements in the PDs well.

Impulse responses showed that shocks in macroeconomic determinants would have created fluctuations in short and medium term when most of the relationships would have disappeared in long term. According to the impulse responses analysis and variance decomposition, the most important relationships to determining the value of the PD of all sectors, if medium term is considered (because in long term, the relationships seem to not exist), are possible positive relationships between the PD and indebtness of the economy, and the PD and real exchange rate, and a negative relationship between the PD and GDP growth if the expectations of the movements correspond to the movements in medium term. Analyses also revealed a long term positive relationship between the PD and inflation, thus inflation might permanently increase the PD if a positive shock in inflation occurred. This relationship though does not follow the expectations.

Impulse responses of sector of non-financial corporations performed almost the same results as the PD of all sectors. The only difference is inflation which seems to be connected with the PD negatively in long term. Hence, higher inflation might help to corporations dealing with their debts according to the analysis which was initially expected.

Sector of households seems to react more slowly and smoother according to the analysis as the relationships between variables are the same as indicated by results of all sectors data. On the other hand, the reaction of the PD of households lacks the initial jump in other direction. The sector of households is the only sector where all of the macroeconomic determinants might be significant. The analysis also suggests that there are no long term relationships except of inflation for which a positive relationship was revealed. In short and medium term, there might be positive relationships between the PD and indebtness, the PD and real exchange rate, and the PD and unemployment, and negative ones between the PD and GDP growth and the PD and interest rates. The

results are not in contradiction to initial expectations except of the indicated relationship between the PD and the interest rates. This might be also result of decrease in interest rates during the crisis when the PD might be higher.
The sector of non-residents performs quite different results than the other sectors. First, it seems to almost not react to shock in macroeconomic variables except of inflation. Moreover, this relationship does not seem to be long term, although the development of the shock reaction might suggest a positive medium term relationship between the PD and inflation.
Finally, the FAVAR model has the best results of forecasting (Table 5.16) in case when the all sectors are modelled together as the forecasting errors are smallest. According to forecasting errors, model might also perform well in case of the sectors of non-financial corporations and households when fitting the evolution of the PD of non-residents might be slightly harder.

Table 5.16: Forecast Measures Applied on Six Periods ahead Forecasts of FAVAR Model Divided by Sectors

Probability of Default	MAPE	RMSE	MFE
All Sectors	13.1517%	0.0017	-0.0014
Corporations	23.1720%	0.0040	0.0038
Households	18.5921%	0.0018	-0.0016
Non-Residents	39.8628%	0.0065	-0.0030

Source: Author's Computations.

Chapter 6

Conclusion

In line with the recent literature, thesis aims to assess relationship between probability of default and macroeconomic determinants. First, credit risk models with macroeconomic variables are listed and according to the data, two models (FAVAR and one latent factor models) are chosen to model probability of default. The monthly data are provided by CNB database covering period 01/2002–06/2013.

Initial ARIMA modelling showed a possible explanation of the PDs by its own values and significance of one, three and four months lagged values, but the assumptions are not fulfilled. Therefore, there might be a bias in estimated coefficients and the results might not be reliable. The advance modelling techniques (Latent factor model and FAVAR) has to be employed. Employing the FAVAR model, it also indicated that the dependence of the PD in its past values might be short or medium term only.

Models performed quite well results in forecasting the future values of the PDs, especially development of the sector of households and the PD including all sectors was fitted very well according to the forecasting measures. The results also indicates that the latent factor model might be slightly better for modelling the sectors of households and non-residents when the FAVAR model might be rather employed in case of PD of whole sectors and the PD of non-financial corporations. Overall, the hypothesis that latent factor model might be better model than FAVAR for PD modelling is not rejected.

Both models suggest some importance of macroeconomic variables in determining PD but the dynamic model – FAVAR – does not indicate any long term relationship between PD and macroeconomic variables except of inflation. Significant relationship between GDP growth and the PD is not confirmed in sector of non-residents by both models. Moreover, the PD sector of households is not indicated to be dependent on GDP growth by latent factor models as well. In all other cases, both models (if medium term is assumed) found negative relationship between the PD and GDP growth. This expected negative relationship is in line with Simons & Rolwes (2009) using data from Netherland, Jiménez & Mencía (2009) employing Mexican data, Jakubík (2006) applying Finnish data, Jakubík (2007), Jakubík & Schmieder (2008) with Czech data, or Fiori *et al.* (2009) on Italian data. In a long term, the negative relationship should disappear according to FAVAR model.

Interest rate was positive significant determinant of the PD of households when FAVAR model was employed and medium term reaction assumed only. In the latent model, the interest rate has to be excluded due to a correlation with unemployment. As unemployment might be very significant determinant of all the PDs, some negative relationship between the PDs and interest rate might exist, but it cannot be confirmed. Thus, it cannot be concluded if the interest rate is significant determinant of the PDs in Czech economy.

Exchange rate is not also examined when the latent factor model is estimated as the correlation with other macroeconomic variables is detected. The FAVAR model resulted in positive relationship between the PDs and exchange rate if the medium term reactions are assumed. Similar relationship was revealed for example by Simons & Rolwes (2009), or Jakubík & Schmieder (2008). This relationship, though, does not hold for the sector of non-residents as the exchange rate is not indicated to be significant determinant of its PD.

Inflation was the only variable which might have a long term effect according to the FAVAR model. A positive long term relationship is indicated by the FAVAR model for all sectors except of the sector of non-financial corporations as the relationship reveal between the PD of non-financial corporations and inflations is negative. Results of latent model estimation suggests similar to FAVAR results that inflation positively influences the PD of households. On the other hand, a different sign is predicted by latent model in case of sector of non-residents. Therefore, the analysis cannot determine exact direction of the effect of inflation on the PD of non-residents.

Unemployment is indicated to be the most important determinant of the all PDs according to the latent factor model estimation. Nevertheless, the FAVAR model suggests that the unemployment might be important only when the PD of households is modelled. Both models indicate expected positive relationship between those variables as for example Hamerle *et al.* (2003) or Jakubík &

Schmieder (2008) found.
Indebtness is also excluded from latent factor model because of its correlation with GDP growth. As the GDP growth might be significant determinant of the PDs of all sectors modelled together, sector of non-financial corporations, and sector of households, there might be also positive influence of indebtness on the PDs as well, although, it cannot be confirmed. Assuming a medium term, FAVAR model indicates positive relationship between indebtness and PDs of all sectors except of non-residents. Moreover, this positive relationship is in line with Jakubík & Schmieder (2008) and Virolainen (2004). In a long run, FAVAR model suggests that the relationship should disappear.
One latent factor model also tests an influence of the financial crisis on probability of default. Results indicate that the financial crisis should significantly influence all modelled PDs except the sector of households. The relationship between the crisis and PDs is suggested to be positive. Therefore, the hypothesis that crisis might have shifted the PDs up in September 2008 is also not rejected.
Further research should be done especially in the area of write-off rate estimation as a variable rate would better reflect true probability of default. Moreover, multi latent factors model might be also employed and compared to the results of one latent factor model to assess if the multi factor model might be more appropriate and if it might be better input for credit risk modelling.

Bibliography

Alessandri, P., P. Gai, S. Kapadia, N. Mora, & C. Puhr (2009): "Towards a Framework for Quantifying Systemic Stability." *International Journal of Central Banking* **5(3)**: pp. 47–81.

Ali, A. & K. Daly (2010): "Macroeconomic determinants of credit risk: Recent evidence from a crosscountry study." *International Review of Financial Analysis* **19**: pp. 165–171.

Basel Committee on Banking Supervision (2001): "The New Basel Capital Accord."

Basel Committee on Banking Supervision (2004): "International Convergence of Capital Measurement and Capital Standards."

Bernanke, B. S., J. Boivin, & P. Eliasz (2004): "Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach." *National Bureau of Economic Research Working Paper 10220* .

Black, F. & M. Scholes (1973): "The Pricing of Options and Corporate Liabilities." *The journal of political economy* pp. 637–654.

Chan-Lau, J. A. (2006): "Fundamentals-Based Estimation of Default Probabilities: A Survey." *IMF Working Paper WP/06/149* .

Chava, S., C. Stefanescu, & S. Turnbull (2011): "Modeling the Loss Distribution." *Management Science* **57(7)**: pp. 1267–1287.

Czech National Bank (2013): "FINANCIAL STABILITY REPORT 2012/2013."

Fernald, J., M. M. Spiegel, & E. Swanson (2013): "Monetary and Fiscal Policy Effectiveness in China: Evidence from a FAVAR Model."

FIORI, R., A. FOGLIA, & S. IANNOTTI (2009): "Beyond macroeconomic risk: the role of contagion in the Italian corporate default correlation." *No. 2. Bocconi University CAREFIN Working Paper* .

FUNGÁČOVÁ, Z. & P. JAKUBÍK (2012): "Bank stress tests as an information device for emerging markets: The case of Russia." *IES Working Paper No. 4/2012* .

GASHA, J. G. & R. A. MORALES (2004): "Identifying Threshold Effects in Credit Risk Stress Testing." *IMF Working Paper WP/04/150* .

GERŠL, A. & J. SEIDLER (2010): "Stress Test Verification as Part of an Advanced Stress-Testing Framework." In "FINANCIAL STABILITY REPORT 2009 / 2010," Czech National Bank.

GREEN, W. H. (2012): *Econometric Analysis*. Pearson Education Limited, 7th edition.

GUPTA, R., M. JURGILAS, & A. KABUNDI (2010): "The effect of monetary policy on real house price growth in South Africa: A factor-augmented vector autoregression (FAVAR) approach." *Economic modelling* **27(1)**: pp. 315–323.

HAMERLE, A., A. DARTSCH, R. JOBST, & P. KILIAN (2011): "Integrating macroeconomic risk factors into credit portfolio models." *The Journal of Risk Model Validation* **5(2)**: p. 3–24.

HAMERLE, A., T. LIEBIG, & D. RÖSCH (2003): "Credit Risk Factor Modeling and the Basel II IRB Approach." *Discussion Paper Series 2: Banking and Financial Supervision No 02/2003* .

HAMERLE, A., T. LIEBIG, & H. SCHEULE (2004): "Forecasting Credit Portfolio Risk." *Discussion Paper Series 2: Banking and Financial Supervision No 01/2004* .

IVANOV, V. & L. KILIAN (2005): "A Practitioner's Guide to Lag Order Selection For VAR Impulse Response Analysis." *Studies in Nonlinear Dynamics & Econometrics* **9(1)**.

JAKUBÍK, P. (2006): "Does Credit Risk Vary with Economic Cycles? The Case of Finland." *IES Working Paper 11/2006* .

Jakubík, P. (2007): "Macroeconomic Environment and Credit Risk." *Czech Journal of Economics and Finance* **57(1-2)**: pp. 60–78.

Jakubík, P. & J. Heřmánek (2006): "Credit Risk and Stress Testing of the Banking Sector in the Czech Republic." In "FINANCIAL STABILITY REPORT," Czech National Bank.

Jakubík, P. & C. Schmieder (2008): "Stress Testing Credit Risk: Is the Czech Republic Different from Germany?" *Czech National Bank Working Papers* **9**.

Jiménez, G. & J. Mencía (2009): "Modelling the distribution of credit losses with observable and latent factors." *Journal of Empirical Finance* **16**: p. 235–253.

Jolliffe, I. T. (2002): *Principal Component Analysis.* Springer, 2nd edition.

Kennedy, P. (2003): *A Guide to Econometrics.* MIT press, 5th edition.

Koopman, S. J. & A. Lucas (2003): "Business and Default Cycles for Credit Risk." *Tinbergen Institute Discussion Paper No. 03-062/2* .

Lowe, P. (2002): "Credit risk measurement and procyclicality." *BIS Working Papers No 116* .

Luetkepohl, H. (2011): "Vector Autoregressive Models." *EUI Working Papers ECO 2011/30* .

Marcucci, J. & M. Quagliariello (2008): "Is bank portfolio riskiness procyclical?: Evidence from Italy using a vector autoregression." *Journal of International Financial Markets, Institutions and Money* **18(1)**: pp. 46–63.

Marcucci, J. & M. Quagliariello (2009): "Asymmetric effects of the business cycle on bank credit risk." *Journal of Banking & Finance* **33 (9)**: p. 1624–1635.

McKinley, S. & M. Levine (1998): "Cubic spline interpolation." *College of the Redwoods* **45**: pp. 1049–1060.

Merton, R. C. (1974): "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *The Journal of Finance* **29(2)**: pp. 449–470.

Pederzoli, C. & C. Torricelli (2005): "Capital requirements and business cycle regimes: forward-looking modelling of default probabilities." *Journal of Banking & Finance* **29(12)**: pp. 3121–3140.

Rösch, D. (2005): "An empirical comparison of default risk forecasts from alternative credit rating philosophies." *International Journal of Forecasting* **21 (1)**: pp. 37–51.

Schuermann, T. & S. G. Hanson (2004): "Estimating probabilities of default." *Staff Report, Federal Reserve Bank of New York, No. 190* .

Simons, D. & F. Rolwes (2009): "Macroeconomic Default Modeling and Stress Testing." *International Journal of Central Banking* **5(3)**: pp. 177–204.

Stein, R. M. (2006): "Are the Probabilities Right? Dependent Defaults and the Number of Observations Required to Test for Default Rate Accuracy." *Journal of Investment Management* **4(2)**: pp. 61–71.

Tchaidze, R. & A.-M. Westin (2010): "Czech Republic: Selected Issues Paper." In "IMF Country Report No. 10/59," International Monetary Fund.

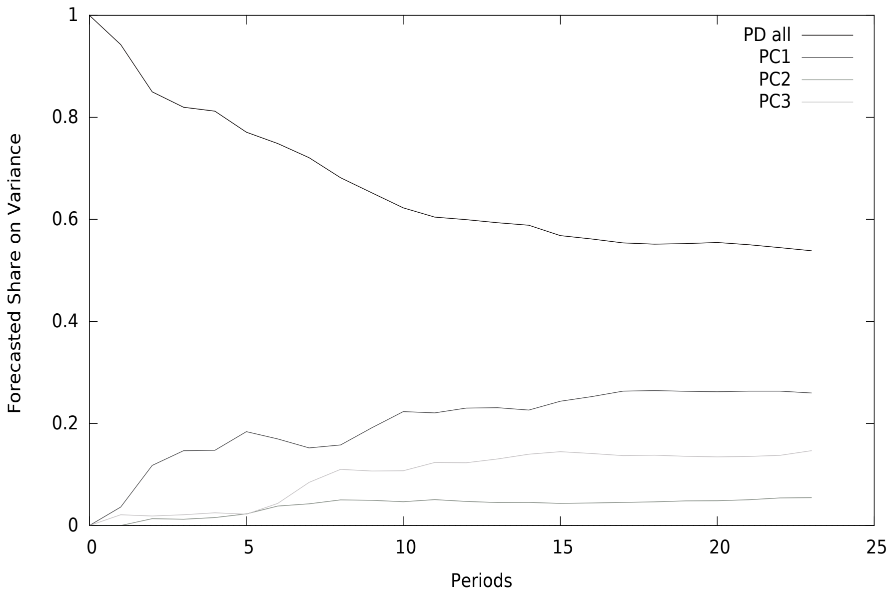
Virolainen, K. (2004): "Macro stress testing with a macroeconomic credit risk model for Finland." *Bank of Finland Discussion Papers 18/2004* .

Appendix A

VAR Impulse Responses and Variance Decomposition

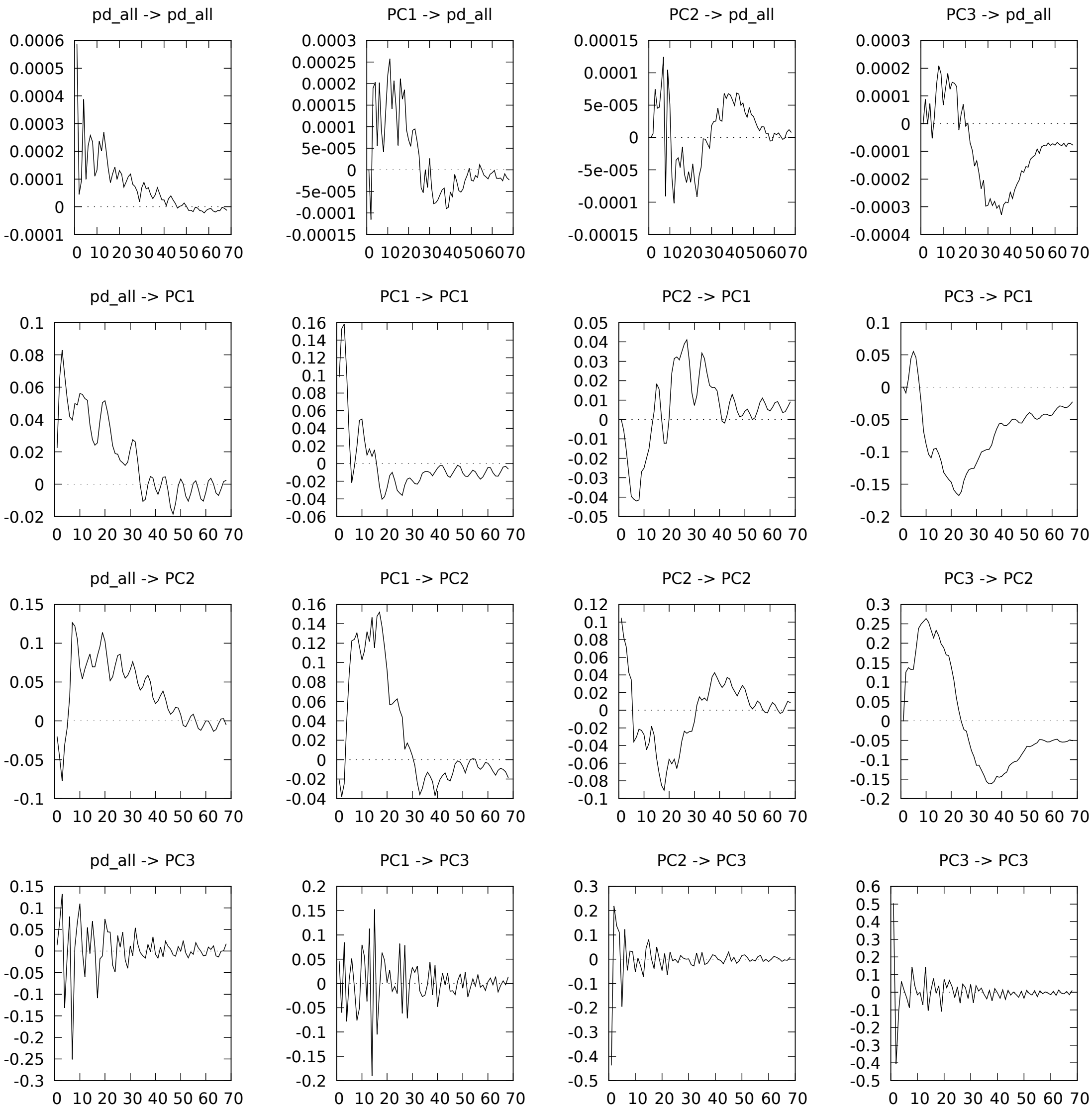
A.1 PD of All Sectors

Figure A.1: Variance Decomposition of FAVAR model – PD of All Sectors



Source: Author's Computations.

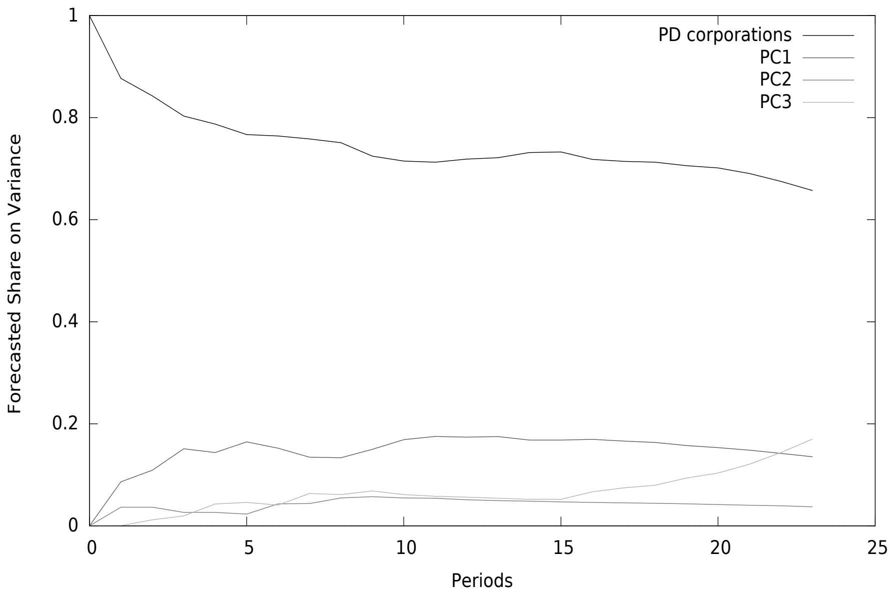
Figure A.2: Impulse Responses of FAVAR model – PD of All Sectors



Source: Author's Computations.

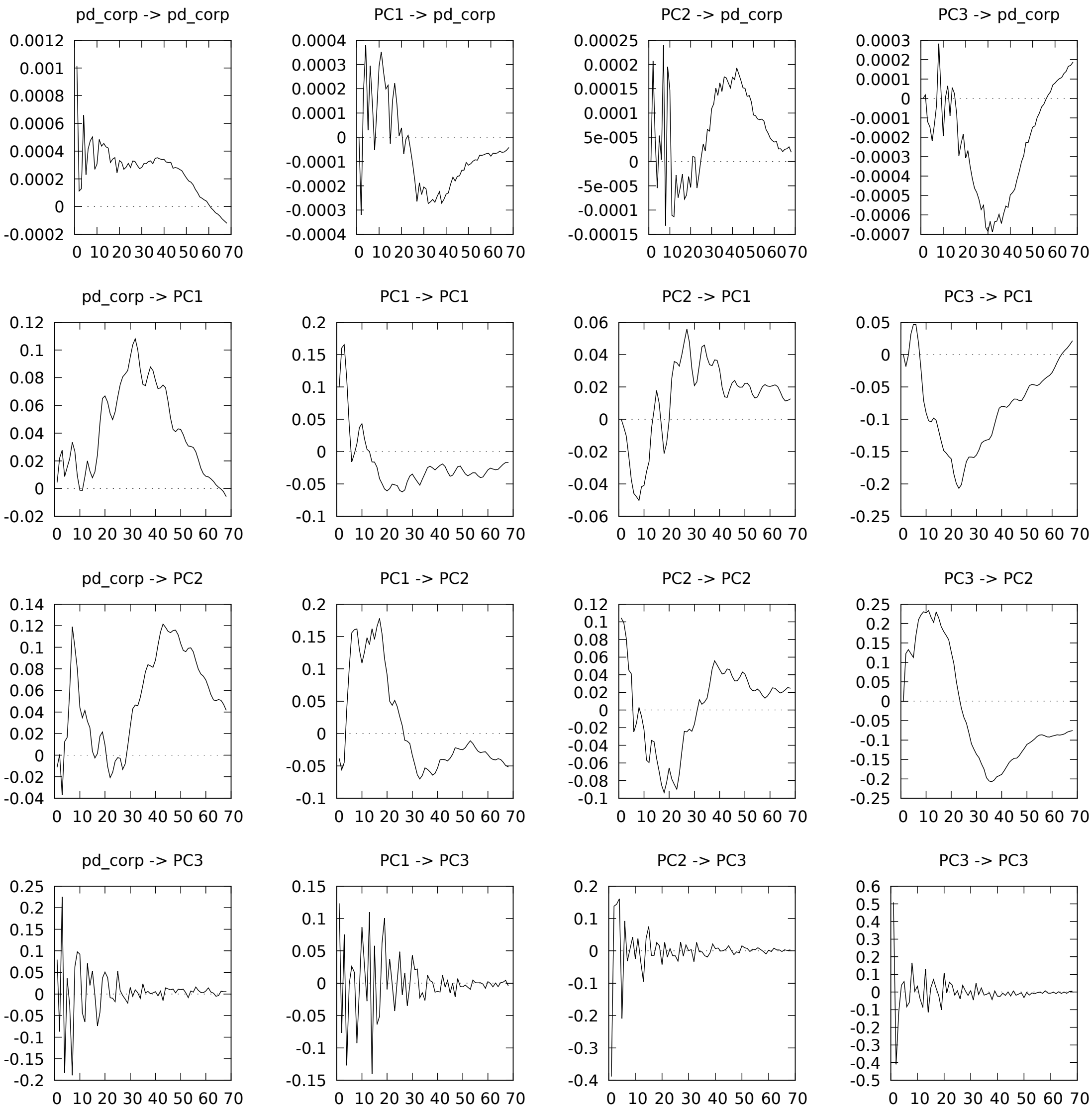
A.2 PD of Non-Financial Corporations

Figure A.3: Variance Decomposition of FAVAR model – PD of Non-Financial Corporations



Source: Author's Computations.

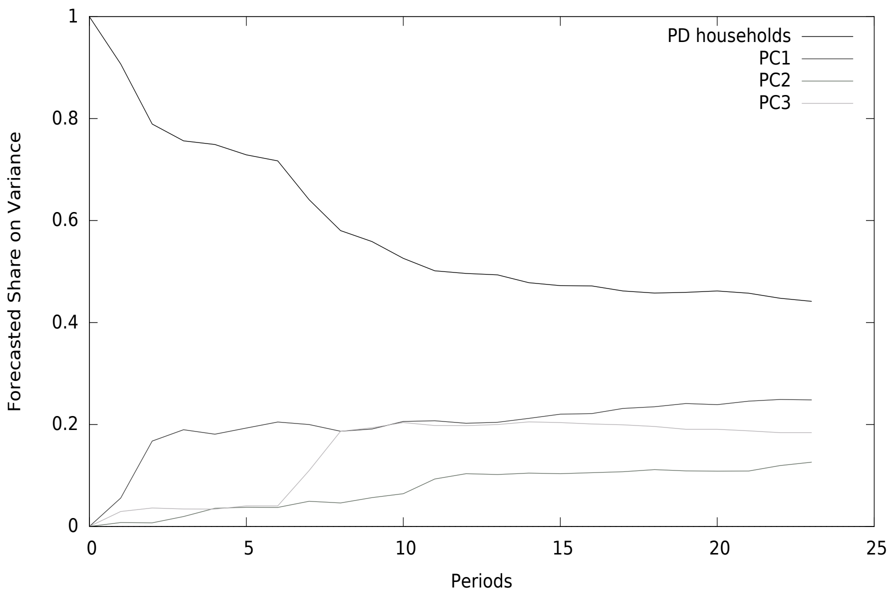
Figure A.4: Impulse Responses of FAVAR model – PD of Non-Financial Corporations



Source: Author's Computations.

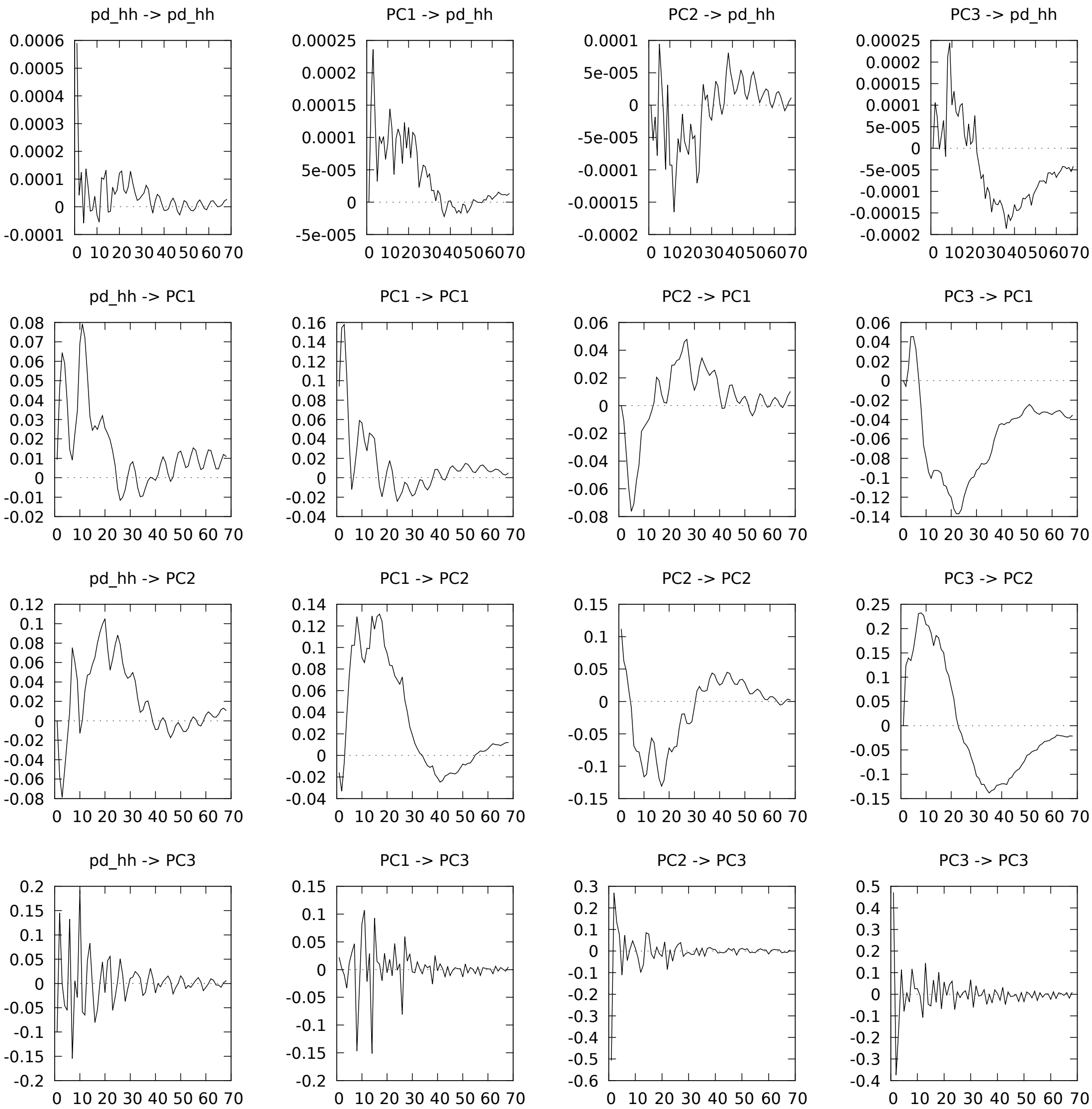
A.3 PD of Households

Figure A.5: Variance Decomposition of FAVAR model – PD of Households



Source: Author's Computations.

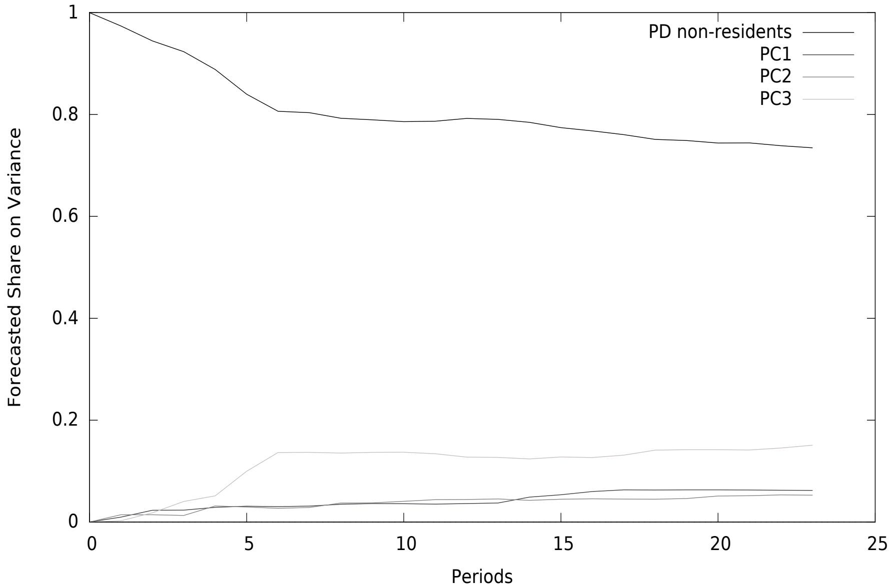
Figure A.6: Impulse Responses of FAVAR model – PD of Households



Source: Author's Computations.

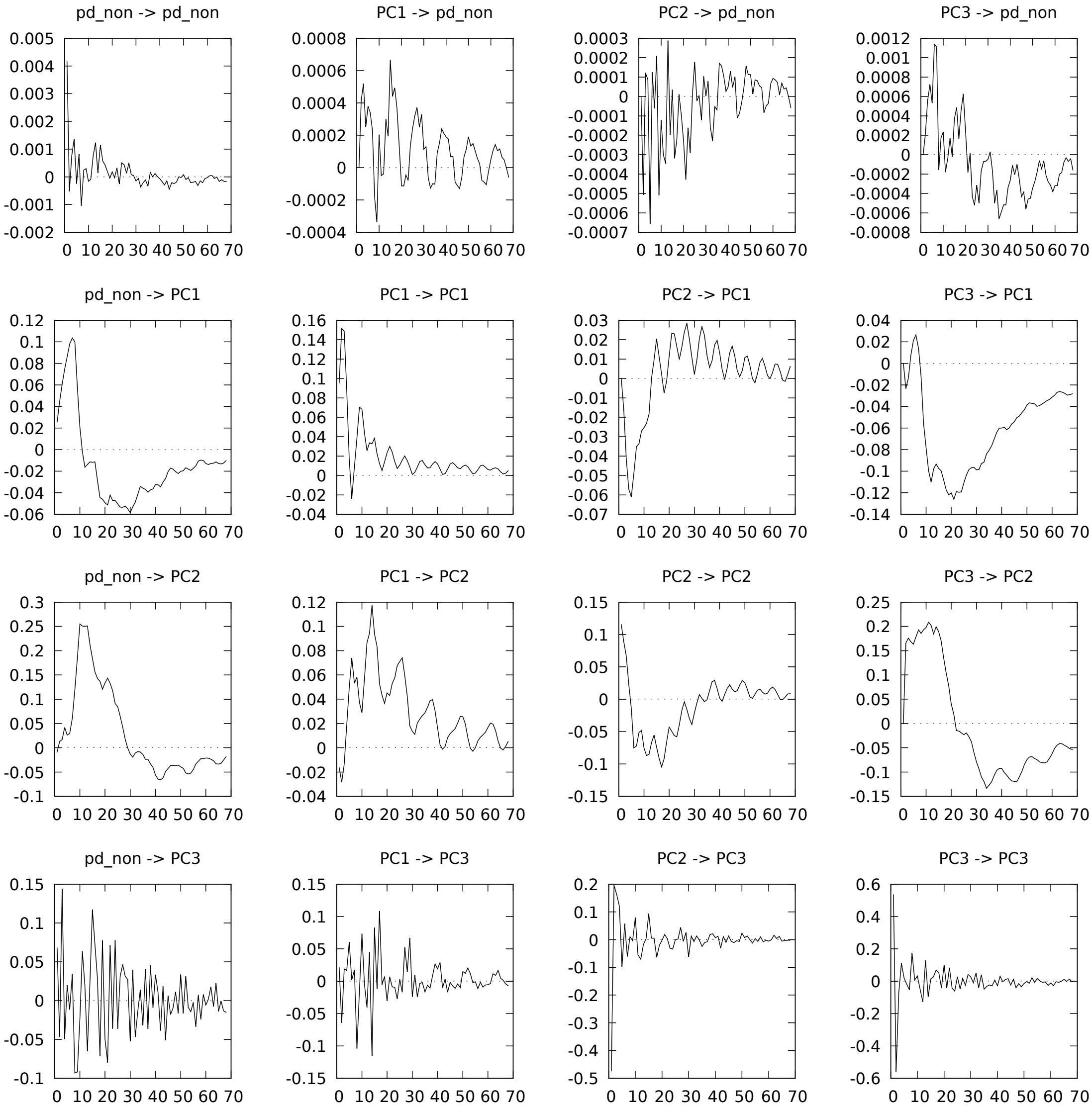
A.4 PD of Non-Residents

Figure A.7: Variance Decomposition of FAVAR model – PD of Non-Residents



Source: Author's Computations.

Figure A.8: Impulse Responses of FAVAR model – PD of Non-Residents



Source: Author's Computations.

Appendix B

Non-Linearity Test

Table B.1: Results of Non-Linearity test – PD of All Sectors

Transition Variable	P-Value	Suggested Model
PD all -1	1.0703e-13	Linear
PD all -2	7.9332e-08	Linear
PD all -3	8.7603e-12	Linear

Source: CNB and Author´s Computations

Table B.2: Results of Non-Linearity test – PD of Non-Financial Corporations

Transition Variable	P-Value	Suggested Model
PD corpor -1	3.1830e-13	Linear
PD corpor -2	1.1523e-09	Linear
PD corpor -3	2.0504e-18	Linear

Source: CNB and Author´s Computations

Table B.3: Results of Non-Linearity test – PD of Households

Transition Variable	P-Value	Suggested Model
PD househ -1	6.4900e-02	Linear
PD househ -2	1.7863e-05	Linear
PD househ -3	7.3229e-02	Linear

Source: CNB and Author´s Computations

Table B.4: Results of Non-Linearity test – PD of Non-Residents

Transition Variable	P-Value	Suggested Model
PD non-res -1	1.3699e-04	Linear
PD non-res -2	1.4921e-23	Linear
PD non-res -3	7.0116e-02	Linear

Source: CNB and Author´s Computations