Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

RIGOROUS THESIS

Volatility Spillovers in New Member States: A Bayesian Model

Author: **Mgr. Radek Janhuba**

Supervisor: **doc. Roman Horváth Ph.D.**

Academic Year: **2012/2013**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature.

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

Prague, September 9, 2013

Signature

Acknowledgments

I would like to thank my supervisor, doc. Roman Horváth Ph.D., for his time, suggestions and valuable comments. I would also like to thank Marek Rusnák from the research department of Czech National Bank for his advices regarding econometric estimation.

Abstract

Volatility spillovers in stock markets have become an important phenomenon, especially in times of crises. Mechanisms of shock transmission from one market to another are important for the international portfolio diversification. Our thesis examines impulse responses and variance decomposition of main stock indices in emerging Central European markets (Czech Republic, Poland, Slovakia and Hungary) in the period of January 2007 to August 2009. Two models are used: A vector autoregression (VAR) model with constant variance of residuals and a time varying parameter vector autoregression (TVP-VAR) model with a stochastic volatility. Opposingly of other comparable studies, Bayesian methods are used in both models. Our results confirm the presence of volatility spillovers among all markets. Interestingly, we find significant opposite transmission of shocks from Czech Republic to Poland and Hungary, suggesting that investors see the Central European exchanges as separate markets.

Bibliographic Record

Janhuba, R. (2012): Volatility Spillovers in New Member States: A Bayesian Model. Master thesis, Charles University in Prague, Faculty of Social Sciences, Institute of Economic Studies. Supervisor: doc. Roman Horváth Ph.D.

JEL Classification	C11, C32, C58, G01, G11, G14
Keywords	Volatility spillovers, Bayesian VAR, TVP-VAR, Stock market

Author's e-mail	`janhuba@gmail.com`
Supervisor's e-mail	`roman.horvath@gmail.com`

Abstrakt

Přelévání volatility akciového trhu se zejména v časech krize stalo důležitým fenoménem. Mechanismy přenosu šoků z jednoho trhu do druhého jsou důležité pro diverzifikaci portfolia v mezinárodním měřítku. Naše diplomová práce zkoumá impulsní odezvy a dekompozici rozptylu čtyř hlavních akciových indexů rozvíjejících se trhů ve střední Evropě (Česká republika, Polsko, Slovensko a Maďarsko) v období od ledna 2007 do srpna 2009. V práci jsou použity dva modely: vektorová autoregrese (VAR) s konstantním rozptylem reziduí a vektorová autoregrese s časově rozdílnými parametry (TVP-VAR) se stochastickou volatilitou. Na rozdíl od jiných porovnatelných studií jsou v obou modelech použity Bayesovké metody. Naše výsledky potvrzují přítomnost přelévání volatility ve všech trzích. Zajímavým zjištěním je nalezení opačného přenosu šoků z České republiky do Polska a Maďarska, což naznačuje, že investoři vidí středoevropské burzy jako oddělené trhy.

Bibliografická evidence

Janhuba, R. (2012): Přelévání volatility v nově členských státech Evropské unie: Bayesovský model. Diplomová práce, Univerzita Karlova v Praze, Fakulta sociálních věd, Institut ekonomických studií. Vedoucí práce: doc. Roman Horváth Ph.D.

Klasifikace JEL	C11, C32, C58, G01, G11, G14
Klíčová slova	Přelévání volatility, Bayesovský VAR, TVP-VAR, Akciový trh

E-mail autora	`janhuba@gmail.com`
E-mail vedoucího práce	`roman.horvath@gmail.com`

Contents

List of Tables

List of Figures

Acronyms

ARCH Autoregressive conditional heteroscedasticity

BVAR Bayesian vector autoregression

CE Central Europe

EHM Efficient market hypothesis

IMW Independent Minnesota Wishart

INW Independent Normal Wishart

IPD Impulse Performance Diagnostics

p.d.f. probability density function

TVP-VAR Time varying parameter vector autoregression

VAR Vector autoregression

SI Spillover Index

SV Stochastic volatility

Chapter 1

Introduction

Together with the growing technology potential and corresponding stock market growth, a numerous research on integration and volatility spillovers of these markets has been performed lately. Such research included advanced as well as emerging economies. However, studies concentrating on integration and volatility spillovers in New Member States have usually examined the relationship of these stock markets to some advanced markets, e.g. German or British. This thesis examines the direct relationship among these stock markets.

We focus our attention to the so called Visegrad countries. Three out of these countries, Czech Republic, Hungary and Poland are the largest new member states that joined EU in 2004. They represent growing stock markets and are often considered the most developed economies from the 2004 accession. Nevertheless, even though these economies developed very much since early 1990's, their stock markets still have not achieved liquidity and levels of market capitalisation that would be comparable to Western European or main world's stock markets. As a result, possible gains from international portfolio diversification into these countries arise (Gilmore & McManus 2002).

The particular research interest lies in impulse response functions and variance decomposition of sample indices as their volatility can be seen as a proxy for their risk (Scheicher 2001). The analysis of impulse responses will reveal information about the transmission of shocks from one country to another. The variance decomposition obtained from impulse responses will show how much of volatility in each country is driven on its own and how much is transmitted from the other countries.

The data are examined using two different vector autoregression (VAR) models. Firstly, we provide a basic homoscedastic vector autoregression. The second model generalizes our first model by allowing its coefficients and volatility of residuals to vary in time. Both models are calculated in Bayesian framework, which is one of the main contributions of our thesis.

Earlier version of this thesis was submitted as a master thesis at the Institute of Economic Studies of the Faculty of Social Sciences at the Charles University in Prague. Changes were made in chapter three in order to refine the used terminology.

The remainder of this thesis is structured as follows. Chapter two provides overview of theory and literature associated with topics of this thesis and chapter three presents basic terms and tools used in Bayesian econometrics. Chapter four presents data and methodology used in the empirical estimation, results of which are presented in chapter five. Chapter six discusses robustness of our results, chapter seven concludes and suggests ideas for future research.

Chapter 2

Theory and Literature Review

2.1 Efficient market hypothesis

Efficient market hypothesis (EMH) is by far the most important concept that has been used in modern finance. In fact, Frankfurter & McGoun (1999) state that "many equate what is called modern finance with the EMH". According to EMH, all markets move in an efficient manner which implies an impossibility of abnormal returns, because any news is immediately negated by the rational behavior of investors. There are three forms of EMH. The weak form only considers historical information, the semi-strong applies for all publicly available information and the strong form includes even privately available information. An interested reader is advised to see Fama (1970) for a detailed overview of the three concepts.

Even though many have tried, up to a current state no one has come with a theory that would generally outperform the EMH (Fama 1998), however, many have shown that EMH does not truly reflect the actual behavior of financial markets. It is beyond the scope of this thesis to provide an overview of all such demonstrations,[1] instead of it we merely state that the sole existence of volatility spillovers provides an example of market inefficiency (Wei-Chong *et al.* 2011).

2.2 Portfolio diversification

A generally known fact is that investors tend to diversify their portfolios in order to reduce their risk. However, many have found that with the growth of

[1] Various challenges to EMH are described in Novak (2008).

technology and corresponding interdependencies, benefits of portfolio diversification in developed markets declined. On the other hand, this did not apply for emerging markets where gains from portfolio diversifications would still exist (Gilmore & McManus 2002).

Though the early emerging market research concentrated mainly on Asian and Latin American countries, with the transition of Central European economies from communist regimes these markets became interesting as well (Gilmore & McManus 2002).

Several studies such as Scheicher (2001), Gilmore & McManus (2002) and Égert & Kočenda (2011) find no evidence of causal relationship of Central European markets, which can be seen as a proof that CE markets were at least at some point interesting in terms of portfolio diversification for investors from developed markets. On the other hand, results of VAR model variance decomposition by Chelley-Steeley (2005) find presence of integration in Central European markets. Results of our analysis will reveal whether investors make differences among particular Central European markets, which would mean existence of additional gains from diversifying portfolios inside of CE markets.

2.3 Shock transmission and volatility spillovers

Cappiello *et al.* (2006) use the regression quantile method to find that the Czech Republic, Hungary and Poland exhibit strong comovements among themselves. Although our research question does not include interdependency of returns, it is a clue that volatility spillovers should exist.

Models assessing volatility spillovers in Central European countries found evidence that such spillovers do exist. Scheicher (2001) found out that shocks in Hungarian market spill over to the Czech market, which spills over to the Polish market.[2] Kasch-Haroutounian & Price (2001) analyze main stock indices of Czech, Polish, Slovak and Hungarian markets and find significant volatility spillovers from Hungarian to Polish market during the 1990s.

Fedorova & Saleem (2010) analyse markets of Czech Republic, Poland, Hungary and Russia in the period from 1995 to 2008. They find an existence of bidirectional shock transmissions for pairs Czech Republic & Poland and Czech Republic & Hungary, but only a one-directional relationship of Poland & Hungary. To the contrary, they find the exactly opposing result for volatility spill-

[2]Sheicher's results hold mainly for returns, but several volatility coefficients are also significant

overs, which are found to be bidirectional for Poland & Hungary, but Czech Republic is dominated by Hungary and at the same time dominates the Polish market.

2.4 TVP-VARs

The advantage of time-varying parameter VAR (TVP-VAR) models lies in estimating different coefficients for each time unit of the sample. As far as the TVP-VAR work is taken generally, researchers have mostly concentrated on various macroeconomic variables, such as relationship of inflation and unemployment (Cogley & Sargent 2001), general monetary policy (Canova & Gambetti (2009), Cogley & Sargent (2005), Koop *et al.* (2009)) or relationship of output and exchange rates in a single country (Mumtaz & Sunder-Plassmann 2010). Canova & Ciccarelli (2006) observe shock transmissions in G-7 countries and Baumeister *et al.* (2008) examine the dynamic effects of liquidity shocks on economic activity, asset prices and infation in Euro area.

Unfortunately, the useful property of estimating huge number of parameters comes with a price of being very demanding in terms of needed computational power. Because of this, only a scarce research has been conducted on financial data. Such research includes Kumar (2010) who runs several models examining the daily exchange rates of Indian currency and finds out that the TVP-VAR model consistently outperforms simple VAR and ARIMA models.

Ito & Noda (2012) run the TVP-VAR model for stock market indices. Specifically, they use impulse responses of a model with Japanese and U.S. markets to find out that stock market linkages and signs of market efficiency do vary in time. However, the dataset of Ito & Noda (2012) contains monthly returns, which means that they lose information about intra-monthly behavior of indices. Our model tries to estimate the TVP-VAR model on a daily stock market data.

Up to our knowledge, only several studies have been conducted on daily stock market data. Sugihara (2010) examines volatility spillovers among European, Japanese and U.S. share and option prices. Triantafyllopoulos (2011) runs a TVP-VAR model for explaining daily stock prices of IBM and Microsoft in the U.S. market.

Chapter 3

Brief Intro to Bayesian Econometrics

The purpose of this chapter is twofold. Firstly, it provides an introduction to the area of Bayesian econometrics for the readers that are unfamiliar with this field. Secondly, it defines some notations and provides definitions used in the remainder of this thesis. It is important to stress that this chapter is by no means a complete guide to the wide field that Bayesian econometrics is. Readers interested in this topic are advised to go through some introductory Bayesian book. A brief and non-mathematical introduction can be found in Koop (2003), somewhat more rigorous and generalized approach is in Dorfman (1997). For more technical analysis including empirical solutions of many methods see Koop *et al.* (2007). Where not stated otherwise, the vast majority of information contained in this section is based on these three books.

The well-known classical, sometimes called *frequentist* econometrics, views parameters of interest as true, unobservable values, about which one is trying to find estimates that are as close as possible to such true values. The biggest difference of Bayesian econometrics is that it takes these parameters as random variables and is consecutively only interested in their distributional properties.

Even though the Bayesian econometrics started as a field in the 1970's, its methods started to blossom with the development of computer hardware. The reason why such methods have recently become used so extensively is that estimation of advanced models commonly requires computing analytically insolvable multidimensional integrals, hence implementation of Bayesian methods often requires usage of numerical software together with advanced hardware.

3.1 Basic ideas

3.1.1 Bayes theorem

The cornerstone of Bayesian econometrics is the Bayes theorem. For events A and B, the definition of conditional probability implies that

$$P(A, B) = P(A|B)P(B). \tag{3.1}$$

By the symmetry of (3.1) in A and B,

$$P(A, B) = P(B|A)P(A). \tag{3.2}$$

Combining (3.1) and (3.2) together, we obtain the simplest form of Bayesian theorem:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}. \tag{3.3}$$

Recalling that Bayesian econometrics views parameters of the model as random variables, it is possible to use conditional probability densities of parameters Θ and data Y to derive analogical version of the Bayes theorem in the form

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}, \tag{3.4}$$

where θ is the value of model parameters we want to estimate about or using the available data y, and $p(\cdot)$ are probability density functions.

In Bayesian framework, $p(\theta|y)$ is of fundamental interest as it directly addresses the question "What do we know about θ given the data y?" (Koop 2003). Using the fact that $p(y)$ does not depend on θ, the term in the denominator can be ignored, which allows us to write

$$p(\theta|y) \propto p(y|\theta)p(\theta) \tag{3.5}$$

Function $p(\theta|y)$ is called the *posterior density* and is used in various methods to establish results of the analysis.[1] Function $p(y|\theta)$ is called the *likelihood function* and function $p(\theta)$ is called the *prior density*. The resulting relationship is sometimes reffered as "posterior is proportional to likelihood times prior" (Koop 2003). The main idea behind this expression is that the knowledge about

[1] Perhaps the most common approach is to evaluate the posterior density to obtain values of θ.

a parameter after seeing the data is a combination of some prior knowledge (independent of the data) and the likelihood function (specifying the distribution of data given the parameters).

3.1.2 Likelihood

The likelihood function $p(y|\theta)$ is sometimes called 'the data generating process'. It specifies the distribution of the data conditional on the parameter values. For example, if errors of data are normally distributed (see the illustrative model in section 3.2), the likelihood function will be the density of normal distribution.

3.1.3 Prior

A prior $p(\theta)$ is a probability function which reflects a set of beliefs that the researcher has about θ before seeing the data. The choice for the researcher is free, however, there are some conventional rules that should be followed while selecting a prior. For example, the prior should not be so centered that it would not allow contribution of the data for updating beliefs about θ (Dorfman 1997).

There are several ways how to divide existing priors according to their characteristics. The first distinction is into *informative* and *noninformative* priors. A noninformative prior does not express any particular beliefs about θ, it simply diffuses all possible information among all possible variants. An example can be a prior in form $p(\theta) = \frac{1}{\sigma}$ in case of linear regression model - such a prior is called *Jeffrey's prior*. On the other hand, an informative prior can restrict some parameters into a range. For example, in a supply and demand equation one can restrict the parameters of θ to positive or negative values according to a set of standard economic assumptions. Alternatively, a normal distribution with chosen exogenous parameters can be specified for the prior.

Another distinction of priors is to *proper* and *improper* priors. A proper prior is such that its probability density integrates to unity. Accordingly, the probability density of improper prior does not integrate to unity, but to some other value, commonly infinity. On the priors from the previous paragraph we can illustrate a common property of priors: that informative priors are often proper, and noninformative priors are commonly improper.

A *conjugate* prior is such that leads to the posterior which allows for its analytical analysis. Moreover, *natural conjugate* prior is a one that comes from the same family of distributions as the likelihood and posterior (Koop & Korobilis 2010). Conjugate priors allow for a considerable simplification of the analysis, however, their importance has declined with the rise of speed and capacity of computing techniques.

A *hierarchical* prior is a prior that depends on some other parameters which are themself calculated in a Bayesian way using a prior on their own. Parameters of such higher prior are called hyperparameters. Hierarchical priors are heavily used in many advanced methods, for an illustrative example see section 3.4 on state space modelling.

Theoretically, a Bayesian prior should be independent on the data as it represents the prior beliefs of the researcher before seeing the data. However, there has been a growing extent of so called *data based* priors which take some prior assumptions using the data. The range of possibilities for such prior is virtually unlimited - for example, Ingram & Whiteman (1994) use the results of business cycle theory models as priors in VAR model. Del Negro & Schorfheide (2004) do the same with results of DSGE models. These papers show that even though data based priors violate the independency condition, it is not unlikely that they will perform well in the empirical analysis.

3.2 Illustrative model

For illustration of basic concepts of Bayesian analysis, we will use the linear regression model in form of equation 3.6. Even though standard estimation of such model requires validity of potentially restrictive assumptions, we can do so using a useful property of of Bayesian inference. It can be shown that many econometric models can be transformed by various techniques to the form of linear regression model. The great feature of Bayesian modelling lies in the fact that complicated models can in many cases be estimated by combining techniques from simpler models in a straightforward manner.[2]

Let us follow the demonstration of Koop (2003) and assume that a regression model is described by equation

$$y = X\beta + \varepsilon, \tag{3.6}$$

where $y = (y_1, y_2, \ldots, y_T)'$ is the vector of realizations of a dependent variable, X is a $T \times k$ matrix of explanatory variables, $\beta = (\beta_1, \beta_2, \ldots, \beta_k)'$ is a $k \times 1$ vector of coefficients and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T)'$ is a $T \times 1$ vector of residuals.

[2] A specific example is the inclusion of heteroscedasticity into the models of Koop & Korobilis (2010).

According to the common approach of Bayesian analysis, let us assume that ε is i.i.d. and following a homoscedastic $N(0_N, h^{-1}I_N)$ where 0_N is a vector of zeros, $h = \sigma^{-2}$ and I_N is $N \times N$ identity matrix so the covariance matrix of residuals is $\sigma^2 I_N$. Hence the set of parameters of interest θ takes the form $\theta = (\beta, h)$.

3.2.1 Likelihood

Using the fact that ε follows the multivariate normal distribution, we can write (see (3.31))

$$p(y|\beta,h) = \frac{h^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}}\left\{\exp\left[-\frac{h}{2}(y - X\beta)'(y - X\beta)\right]\right\}. \tag{3.7}$$

As we do not know anything about h and β, we need to approximate it. The most convenient way is to use OLS estimates.[3] Therefore, we have

$$\hat{\beta} = (X'X)^{-1} X'y \tag{3.8}$$

and

$$s^2 = \frac{\left(y - X\hat{\beta}\right)'\left(y - X\hat{\beta}\right)}{N-k}. \tag{3.9}$$

Putting (3.8) and (3.9) into (3.7), it can be shown that the likelihood function transforms to the form

$$p(y|\beta,h) = \frac{1}{(2\pi)^{\frac{N}{2}}}\left\{h^{\frac{k}{2}}\exp\left[-\frac{h}{2}(\beta - \hat{\beta})'X'X(\beta - \hat{\beta})\right]\right\}\left\{h^{\frac{N-k}{2}}\exp\left[-\frac{h(N-k)}{2s^{-2}}\right]\right\}. \tag{3.10}$$

Such a form of likelihood function will be useful later in the analysis. Note that the middle term is the p.d.f. of the multivariate normal distribution (equation 3.31) and the last term can be interpreted as a p.d.f. of Gamma distribution (equation 3.32).

3.2.2 Prior

There are several ways how to select a prior for the linear regression model. Following Koop (2003), we show the *natural conjugate prior* which allows for the analytical examination of the resulting posterior distribution. Due to the

[3]Details about these estimates can be found in many introductory econometric books, e.g. Greene (2002).

form of (3.10), if we set $\nu = N - k$, the natural conjugate prior requires that the prior for h comes from a gamma distribution

$$h \sim G(\underline{s}^{-2}, \underline{\nu})$$

and the prior for β conditional on h comes from the multivariate normal distribution

$$\beta|h \sim N(\underline{\beta}, h^{-1}\underline{V}).$$

Putting together, we have a Normal-Gamma distribution

$$\beta, h \sim NG(\underline{\beta}, \underline{V}, \underline{s}^{-2}, \underline{\nu}). \tag{3.11}$$

The bar under the parameter, $\underline{\bullet}$, means the arbitrarily chosen initial belief about a prior density, therefore, it is a number which is to be chosen freely by the researcher. In the next section we will introduce a bar over the parameter, $\overline{\bullet}$, which will be an updated value of the posterior parameter after the data come in.

3.2.3 Posterior

Following (3.5), the posterior is obtained by multiplication of the likelihood from (3.10) by the probability density $f_{NG}(\beta, h|\underline{\beta}, \underline{V}, \underline{s}^{-2}, \underline{\nu})$ obtained from (3.11). Thanks to the fact that the prior is natural conjugate, we obtain the posterior distribution

$$\beta, h|y \sim NG(\overline{\beta}, \overline{V}, \overline{s}^{-2}, \overline{\nu}), \tag{3.12}$$

where

$$\overline{V} = \left(\underline{V}^{-1} + X'X\right)^{-1}, \tag{3.13}$$

$$\overline{\beta} = \overline{V}\left(\underline{V}^{-1}\underline{\beta} + X'X\hat{\beta}\right), \tag{3.14}$$

$$\overline{\nu} = \underline{\nu} + N \tag{3.15}$$

and $\overline{s}^{-2}$ satisfies the condition

$$\overline{\nu s}^2 = \underline{\nu s}^2 + \nu s^2 + \left(\hat{\beta} - \underline{\beta}\right)' \left[\underline{V} + (X'X)^{-1}\right]^{-1} \left(\hat{\beta} - \underline{\beta}\right). \tag{3.16}$$

Equation 3.12 represents the joint posterior distribution of β and h. It is possible to marginalize out h in order to obtain the marginal distribution of β

without the influence of h. This can be done using the fact that

$$p(\beta|y) = \int p(\beta, h|y)\mathrm{d}h. \tag{3.17}$$

It can be shown that marginalizing out h leads to a multivariate-t distribution

$$\beta|y \sim t(\overline{\beta}, \overline{s}^2\overline{V}, \overline{\nu}), \tag{3.18}$$

therefore (see section 3.6)

$$E(\beta|y) = \overline{\beta}, \tag{3.19}$$

$$var(\beta|y) = \frac{\overline{\nu}\overline{s}^2}{\overline{\nu} - 2}\overline{V}. \tag{3.20}$$

Moreover, as $\beta, h|y \sim NG(\overline{\beta}, \overline{V}, \overline{s}^{-2}, \overline{\nu})$, we have $h|y \sim G(\overline{s}^{-2}, \overline{\nu})$, hence

$$E(h|y) = \overline{s}^{-2}, \tag{3.21}$$

$$var(h|y) = \frac{2s^{-2}}{\overline{\nu}}. \tag{3.22}$$

Equation 3.19 can be interpreted as a weighted average of OLS coefficient $\hat{\beta}$ and prior mean $\underline{\beta}$, with the weights being $X'X$ and $\underline{V}^{-1}$ (Strasky 2010). Thus, Bayesian estimation in this setting combines classical frequentist approach with certain prior beliefs about parameters of interest. One interesting result arises if we set the noninformative prior in the way that $\nu = 0$ and $\underline{V}^{-1} = 0$, as such estimate of β is equal to $\hat{\beta}$ from OLS (Koop 2003). Therefore, one can use Bayesian techniques to obtain results equal to the classical sampling theory approach.

3.2.4 Posterior analysis

While analyzing posterior characteristics, one can use the definition of conditional expected value

$$E\left(g(\theta)|y\right) = \int g(\theta)p(\theta|y)\mathrm{d}\theta \tag{3.23}$$

using the fact that the posterior is a function of θ. Equation 3.23 might seem little too abstract, therefore we show two examples of likely the most common integration in Bayesian inference. Let us assume that one tries to estimate a value of some parameter, denoted by θ_i. To do so, its mean and variance are needed in order to specify the confidence intervals where the value of parameter lies with some chosen probability. Thus, function g has the form of $g(\theta) = \theta_i$ (equation 3.24) and $g(\theta) = \theta_i^2$ (equation 3.25):[4]

$$E(\theta_i|y) = \int \theta_i p(\theta|y)\mathrm{d}\theta \tag{3.24}$$

$$E\left(\theta_i^2|y\right) = \int \theta_i^2 p(\theta|y)\mathrm{d}\theta \tag{3.25}$$

The natural conjugate prior used in this illustrative example is only one of many priors that could be used. As the choice of a prior could considerably affect results of the estimation, *prior sensitivity analysis* is used to test for robustness of results. This analysis consists of repeating the model estimation using several prior values (and, if applicable, several different priors).

3.3 Nonconjugate priors

Natural conjugate priors described in the previous section have a huge advantage that analytical results are available for integrals in (3.23), therefore no posterior simulation is required. However, they also have some undesirable properties that should be kept in mind when a natural conjugate prior is used. For example, usage of a natural conjugate prior in VAR modelling implies that covariances of coefficients of explanatory variables are proportional to each other, a property that might be undesirable (Koop & Korobilis 2010). Consecutively, non-conjucate priors are often used.

As the analytical results for non-conjugate priors are not available, numerical simulations of posterior densities are required. There are several ways how to approach such analysis. This section describes two particular methods, the Monte Carlo integration and Gibbs sampling. For a concise review of these and other posterior simulation methods, see Tanner (1996).

3.3.1 Monte Carlo integration

Monte Carlo integration is a very general class of methods which allows us to estimate $E(g(\theta)|y)$. To illustrate how it works, let us define a new notation

[4] (3.25) is needed because by the definition of conditional variance, $Var(\theta_i|y) = E\left(\theta_i^2|y\right) - [E(\theta_i|y)]^2$.

$\theta^{(s)}$ which marks the s-th draw from $p(\theta|y)$. Moreover, let us define

$$\hat{g}_S = \frac{1}{S}\sum_{s=1}^{S} g\left(\theta^{(s)}\right). \tag{3.26}$$

The law of the large numbers implies that if all $\theta^{(s)}$ are random, $\lim_{S\to\infty} \hat{g}_S = E\left(g(\theta)|y\right)$. In empirical estimation, we can obtain the mean of parameter θ_i by taking random draws from (3.24) and analogously use (3.25) to obtain confidence intervals. These are obtained using the properties of normal distribution as all distributions converge to normal when $S \to \infty$ according to the central limit theorem.

Unfortunately, it is not always possible to take random draws from the probability density $p(\theta|y)$. In such cases Monte Carlo integration cannot be used.

3.3.2 Gibbs sampling

Even though it is often not possible to take draws from probability density $p(\theta|y)$, the conditional distributions of subsets of θ commonly have forms that allow to draw from them. Without loss of generality, let us say that θ can be divided into three blocks θ_1, θ_2 and θ_3. The Gibbs sampling is performed in the following way:

1. Choose the initial values $\theta_i^{(0)}$,
2. Draw $\theta_1^{(1)}$ from $p\left(\theta_1|y,\theta_2^{(0)},\theta_3^{(0)}\right)$,
3. Draw $\theta_2^{(1)}$ from $p\left(\theta_2|y,\theta_1^{(1)},\theta_3^{(0)}\right)$,
4. Draw $\theta_3^{(1)}$ from $p\left(\theta_3|y,\theta_1^{(1)},\theta_2^{(1)}\right)$,
5. Draw $\theta_1^{(2)}$ from $p\left(\theta_1|y,\theta_2^{(1)},\theta_3^{(1)}\right)$,
6. $\vdots$
7. Draw $\theta_3^{(S)}$ from $p\left(\theta_3|y,\theta_1^{(S)},\theta_2^{(S)}\right)$.

We can see that $\theta^{(s)}$ is dependent on $\theta^{(s-1)}$. Such process is generally called a *Markov process*. Therefore, the Gibbs sampler is an example of so called

Markov Chain Monte Carlo procedure, a wide range of algorithms that use such draws.

Even though the dependence of $\theta^{(s)}$ and $\theta^{(s-1)}$ is a violation of the Monte Carlo integration assumptions, it has been proven that with S going to infinity, the value of $\hat{g}(\theta)$ computed from $\theta^{(j)}$'s converges to $E\left(g(\theta)|y\right)$ (Geweke 1999). To assure that the choice of $\theta_i^{(0)}$ does not have influence on the resulting draws of θ_i, S is divided into S_B and S_D, where the S_B marks the number of *burn-in* iterations where draws of θ_i are not stored.

3.4 State space models

State space models have been extensively used in the empirical research with both Bayesian and frequentist approaches. It is a very general and wide class of models that can incorporate many of the well known models such as e.g. ARMA and VAR models. The reason why this section is included here is that understanding basics of state space models is a necessary requirement before continuing towards TVP-VAR models (Koop & Korobilis 2010). It also shows how the MCMC method can be implemented in the framework of state space models.

Following Koop & Korobilis (2010), a general state space model can be written by a set of equations

$$y_t = W_t\delta + Z_t\beta_t + \varepsilon_t, \tag{3.27}$$

$$\beta_{t+1} = \Pi_t\beta_t + u_t, \tag{3.28}$$

where y is an $M \times 1$ vector of dependent variables, W_t is a known $M \times p_0$ matrix of explanatory variables with constant coefficients represented by a $p_0 \times 1$ vector δ, Z_t is a known $M \times k$ matrix of explanatory variables with time varying coefficients represented by a $k \times 1$ time varying vector β_t. Errors ε_t and u_t are independent in time and each other with $\varepsilon_t \sim N(0, \Sigma_t)$, $u_t \sim N(0, Q_t)$, $Cov(\varepsilon_t, u_s) = 0$ for $t, s = 1, \ldots, T$ and finally $Cov(\varepsilon_i, \varepsilon_j) = Cov(u_i, u_j) = 0$ for $i, j = 1, \ldots, T$ and $i \neq j$. We also assume the $k \times k$ matrix Π_t to be known.

Equation 3.27 is called the *measurement* equation and equation 3.28 is called the *state* equation. In Bayesian estimation, it is necessary to implement priors for δ, β_t, Π_t, Q_t and Σ_t. The posterior density after combining these priors with a specific likelihood function will be analytically unobservable so a computer algorithm that will draw from conditional densities is required (see section 3.3).

3.5 Stochastic volatility

The well known fact about financial time series models is that their residuals often vary in time. There are two main approaches to account for heteroscedasticity - the autoregressive conditional heteroscedasticity (ARCH) and the stochastic volatility. As we only use the latter approach in our model, we will not describe ARCH in this thesis. A brief and concise review of its properties can be found in Brooks (2008), alternatively Bauwens *et al.* (1999) provide its detailed Bayesian treatment together with other time series methods. Kim *et al.* (1998) discuss the differences between ARCH and stochastic volatility models in Bayesian framework. Jeantheau (2004) shows an example of a stochastic volatility model that has very similar properties as GARCH(1,1) model.

The principle of stochastic volatility lies in rearranging residuals into the form

$$y_t = \varepsilon_t \exp\left(\frac{h_t}{2}\right), \tag{3.29}$$

where $\varepsilon_t \sim N(0,1)$ and $h_t \sim N\left(0, \sigma_\eta^2\right)$. The volatility component h_t is then modeled as a random walk following

$$h_{t+1} = h_t + \eta. \tag{3.30}$$

Note that equations 3.29 and 3.30 can be seen as a specific class of state space models described above. As such, stochastic volatility can be incorporated into various state space models. The simplest stochastic volatility model illustrated here allows for many extensions. For example, Kim *et al.* (1998) include a coefficient ϕ into equation 3.30, changing its nature from random walk into an AR(1) process.[5] Primiceri (2005) shows an extension to the multivariate framework. His methodology is used later in this monograph.

[5] Kim *et al.* (1998) also develop likelihood inference for stochastic volatility models, which has been widely used in the following empirical work.

3.6 Statistical distributions

This section provides a basic overview of statistical distributions appearing in this monograph. The purpose of this overview is to show the probability density functions, parameters, means and variances of distributions that might be unknown for a reader of this text. More details can be found in a variety of statistical and econometric books, e. g. Koop (2003), Koop *et al.* (2007). A very detailed treatment of these and many more statistical distributions including rigorous proofs and derivations can be found in chapter 3 of Poirier (1995).

Multivariate normal distribution

A random variable y is said to follow multivariate normal distribution $\phi(y|\mu, \Sigma)$ with mean μ and variance Σ, denoted by $y \sim N(\mu, \Sigma)$, if its p.d.f. is

$$\phi(y|\mu, \Sigma) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(y-\mu)^T\Sigma^{-1}(y-\mu)\right), \tag{3.31}$$

where y and μ are k-dimensional vectors and Σ is a $k \times k$ positive definite matrix.

Gamma distribution

A random variable y is said to follow a gamma distribution, denoted by $y \sim G(\alpha, \beta)$, if its p.d.f. is

$$f_\gamma(y|\mu, \nu) = \begin{cases} c_G^{-1} y^{\alpha-1} \exp\left(-\frac{y}{\beta}\right) & \text{if } 0 < y < \infty, \\ 0 & \text{otherwise,} \end{cases} \tag{3.32}$$

where

$$c_\gamma = \beta^\alpha \Gamma(\alpha)$$

is the integrating constant and $\Gamma(\alpha)$ is the gamma function satisfying

$$\int_0^\infty t^{\alpha-1} \exp(-t)\mathrm{dt}.$$

The mean of gamma distribution is $\alpha\beta$ and its variance is $\alpha\beta^2$ (Poirier 1995). The gamma distribution is a generalization of some well known distributions - specifically, if $\alpha = 1$, it is the exponential distribution and if $\beta = 2$, it is a Chisquare distribution. The inverted gamma distribution is also used extensively, meaning that if y has an inverted gamma distribution, then $1/y$ has a gamma distribution.

Normal-Gamma distribution

Let h be a random variable following a gamma distribution $G(m, \nu)$ and y be a random vector following the conditional normal distribution $y|h, \mu, \Sigma \sim N(\mu, h^{-1}\Sigma)$. Then $\theta = (y', h)'$ follows the Normal-Gamma distribution denoted by $\theta \sim NG(\mu, \Sigma, m, \nu)$.

Multivariate-t distribution

A continuous k-dimensional random vector y has a multivariate-t distribution with a mean μ, scale matrix Σ and degrees of freedom ν, denoted by $y \sim t(\mu, \Sigma, \nu)$, if its p.d.f. is denoted by

$$f_t(y|\mu, \Sigma, \nu) = \frac{1}{c_t} |\Sigma|^{-\frac{1}{2}} \left[\nu + (y - \mu)'\Sigma^{-1}(y - \mu)\right]^{-\frac{\nu+k}{2}},$$

where

$$c_t = \frac{\pi^{\frac{k}{2}} \Gamma\left(\frac{\nu}{2}\right)}{\nu^{\frac{\nu}{2}} \Gamma\left(\frac{\nu+k}{2}\right)}.$$

In order for the multivariate-t distribution to have a defined mean and variance, the condition $\nu > 2$ has to be satisfied.[6] In such cases, the mean of the distribution is $E(Y) = \mu$ and its variance is $var(Y) = \frac{\nu}{\nu-2}\Sigma$.

In the univariate case, if we set $\mu = 0$ and $\Sigma = 1$, we have a well-known Student-t distribution with ν degrees of freedom.

Wishart distribution

The Wishart distribution is a multivariate generalization of the Gamma distribution defined above. The $N \times N$ random positive definite symmetric matrix H has a Wishart distribution with a scale matrix A ($N \times N$, known and positive definite) and degrees of freedom ν (positive scalar), denoted by $H \sim W(A, \nu)$, if its p.d.f. is given by

$$f_W(H|A, \nu) = \frac{1}{c_W} |H|^{\frac{\nu-N-1}{2}} |A|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\, tr(A^{-1}H)\right],$$

[6] If $1 < \nu < 2$, the mean exists, but the variance does not.

where

$$c_W = 2^{\frac{\nu N}{2}} \pi^{\frac{N(N-1)}{4}} \prod_{i=1}^{N} \Gamma\left(\frac{\nu + 1 - i}{2}\right).$$

If we denote H_{xy} to be an element of the matrix H in the x-th row and y-th collumn, then for $i, j, k, m = 1, \cdots, N$ the mean of the Wishart distribution is $E(H_{ij}) = \nu A_{ij}$, its variance is $var(H_{ij}) = \nu \left(A_{ij}^2 + A_{ii}A_{jj}\right)$ and the covariance of two distinct elements is $cov(H_{ij}, H_{km}) = \nu \left(A_{ik}A_{jm} + A_{im}A_{jk}\right)$.

Chapter 4

Data and Methodology

In order to capture interdependencies among multiple time series, the vector autoregression (VAR) model was chosen. The lag length of 5 was selected to account for possible dependencies originating from one week prior to the particular observation.

4.1 Data

The stock market data were downloaded from particular stock exchanges' websites. Data were collected for the period of January 2007 to August 2009. This time span was chosen as it includes several periods which differ in terms of market conditions, hence it is likely that properties of the data changed in time and amplitude of shocks changed as well. As the amount of shocks tends to increase in times of crisis, higher volatility of markets during such time is expected.

Figure 4.1 shows the time behavior of the four stock indices during the sample period. The two cut-off dates depicted by dashed lines were chosen arbitrarily based on the data properties. The first cut-off date, January 16 2008, was chosen as it is the date when Czech, Polish and Hungarian markets all fell under the level of the first observation and stayed below this level until the end of sample period.[1] The second cut-off date, March 6 2009, was chosen as it is the first date since September 2008 where all four markets rose in the day immediately following the day where all four of them declined.

We can see that all indices share common properties. In fact, behavior of the

[1] To be precise, this statement holds with the exception of 28 observations in the period of May 2 to June 18 2008 (34 trading days in total).

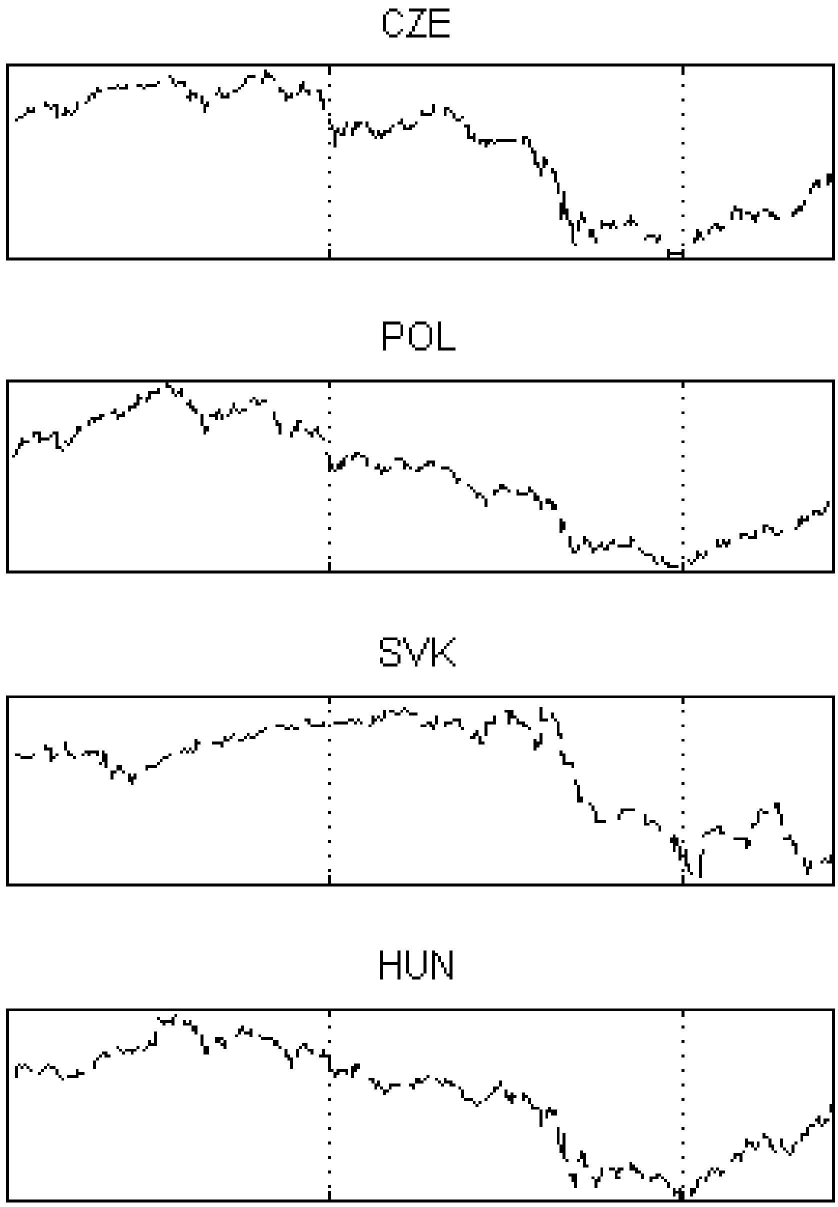

Figure 4.1: Stock market indices, January 2007 to August 2009

Czech, Polish and Hungarian indices is extremely similar. The sample can be broken down into three periods which differ in terms of market behavior and are marked by the vertical lines. The first period represents a *stable* development where stock indices behave without a general trend. Afterwards, the *crisis* comes and markets start to decline. In March 2009 markets *rebound* towards an increasing trend.[2]

Even though the (non)stationarity of the data is irrelevant in Bayesian framework, all series were log-differenced in order to assure comparability with the standard sampling theory research. The second reason to use log-differenced data is to avoid the potential danger of spurious regression present in time series modelling.[3]

Table 4.1: Descriptive statistics of the full sample

	CZE	POL	SVK	HUN
Minimum	-0.1619	-0.0829	-0.0958	-0.1265
Maximum	0.1236	0.0608	0.1188	0.1318
Mean	-0.0005	-0.0004	-0.0004	-0.0004
Std. Dev	0.0226	0.0175	0.0113	0.0220
Variance	0.0005	0.0003	0.0001	0.0005
Skewness	-0.4450	-0.2897	0.3497	-0.0696
Kurtosis	12.4862	4.9852	33.8124	8.9799
Jarque - Bera	2572.0738	121.1766	26913.7294	1013.7143

Note: Jarque-Bera test statistics is significant at any imaginable level of confidence (critical value is 5.9706).

Descriptive statistics of the full sample is presented in table 4.1. Each variable contains 680 observations, which totals to the sample size of 2720. We can see that data from all countries have a structure that is typical for financial time series'. All series except Slovakia are skewed to the left which means that there were relatively more declines than increases (on the other hand, these relatively few increases had a relatively higher magnitude). We can also see that all four series are leptocurtic, which is a very common property of financial

[2]Note that there is a difference in the behavior of Slovak market which contains very low but stable increasing trend from approximately half of the stable period to approximately half of the crisis.

[3]Spurious regression arises when two or more series contain a trend - in such cases one can find significant statistical relationship among variables that do not have any causal relations (see Granger & Newbold (1974)).

data (Brooks 2008).[4] The Jarque-Bera statistics show that all series are highly non-normal.

Table 4.2 shows the descriptive statistics of the three subsamples described above. Period 1 runs from January 2007 to January 15 2008 and contains 264 observations for each time series. Period 2 runs from January 16 2008 to March 5 2009 and contains 291 observations. Period 3 runs from March 6 2009 to the end of August 2009 and contains 125 observations.

Table 4.2: Descriptive statistics of subsamples

	CZE			POL		
	S1	S2	S3	S1	S2	S3
Min.	-0.0567	-0.1619	-0.0644	-0.0631	-0.0829	-0.0514
Max.	0.0274	0.1236	0.0612	0.0446	0.0608	0.0580
Mean	0.0001	-0.0032	0.0048	-0.0002	-0.0026	0.0042
St.d.	0.0110	0.0293	0.0219	0.0135	0.0200	0.0181
Var.	0.0001	0.0009	0.0005	0.0002	0.0004	0.0003
Skew.	-0.8408	-0.2763	-0.0873	-0.4241	-0.3443	0.2337
Kurt.	5.6415	9.5178	3.5987	4.8142	4.5918	3.4771
J-B.	107.8618	518.7920	2.0256	44.1172	36.4740	2.3236
	SVK			HUN		
	S1	S2	S3	S1	S2	S3
Min.	-0.0301	-0.0513	-0.0958	-0.0436	-0.1265	-0.0463
Max.	0.0236	0.0624	0.1188	0.0334	0.1318	0.0640
Mean	0.0003	-0.0013	-0.0001	-0.0000	-0.0032	0.0056
St.d.	0.0064	0.0104	0.0188	0.0117	0.0269	0.0246
Var.	0.0000	0.0001	0.0004	0.0001	0.0007	0.0006
Skew.	-1.1698	-0.1938	0.5931	-0.0874	-0.0650	0.1580
Kurt.	8.7564	12.8716	20.908	4.1236	8.3103	2.3730
J-B.	424.7190	1183.3737	1677.6157	14.2233	342.1167	2.5681

Several properties of subsamples are worth mentioning. Firstly, we can see that means of all series during the crisis are negative, which is an expected property. Similar statement holds for means in the rebound period, where only Slovakia with a mean of -0.0001 reports negative value. Secondly, the data are negatively skewed in the first two periods and positively skewed during the rebound period (with the exception of the Czech Republic), a property which is also expected. The most interesting and very unexpected finding is that,

[4] Leptocurtic distributions are sometimes said to contain 'fat tails'.

except for the Slovak market, the Jarque-Bera test applied to the data from the rebound period does not reject the null hypothesis of normality.

4.2 BVAR

Our first model is described by equation

$$y_t = c + A^1 y_{t-1} + A^2 y_{t-2} + A^3 y_{t-3} + A^4 y_{t-4} + A^5 y_{t-5} + \varepsilon_t \tag{4.1}$$

where y is an M-dimensional vector of examined variables, c is an M-dimensional vector of constants, A^i is an $M \times M$ matrix of coefficients for the i-th lag of examined variables and ε_t is an M-dimensional homoscedastic vector of random errors following $\varepsilon_t \sim N(0, \Sigma)$. If we set

$$\begin{aligned} A &= \begin{bmatrix} c & A^1 & A^2 & A^3 & A^4 & A^5 \end{bmatrix}, \\ X &= \begin{bmatrix} 1 & y_{t-1} & y_{t-2} & y_{t-3} & y_{t-4} & y_{t-5} \end{bmatrix}', \end{aligned} \tag{4.2}$$

equation 4.1 can be rewritten as

$$y_t = X_t A + \varepsilon_t, \tag{4.3}$$

which can be rearranged into the form

$$y_t = Z_t \alpha + \varepsilon_t, \tag{4.4}$$

where

$$Z_t = (I \otimes X_t) \tag{4.5}$$

and $\alpha = vec(A)$. The biggest advantage of such trasformation is that the residuals are normally distributed following $\varepsilon_t \sim N(0, \Sigma \otimes I_T)$, which allows the researcher to break the sampling density $p(y|\alpha, \Sigma)$ into two separate parts (see Koop & Korobilis (2010)).

4.2.1 Independent Normal-Wishart prior

Following Koop & Korobilis (2009), we decided to use two different priors in our estimation of the basic BVAR model. The first prior is called the Independent Normal-Wishart prior. As the Z_t from (4.4) equals to $I \otimes X_t$ and the residual

ε_t has a variance matrix $\Sigma \otimes I_T$, we can combine the multivariate normal distribution for obtaining draws of α with drawing the variance matrix Σ from the Wishart distribution. As the two priors are independent of each other, we use the basic probability rule that

$$p\left(\alpha, \Sigma^{-1}\right) = p\left(\alpha\right) p\left(\Sigma^{-1}\right). \tag{4.6}$$

The prior takes form of

$$\alpha \sim N\left(\underline{\alpha}, \underline{V}_{\alpha}\right), \tag{4.7}$$

$$\Sigma^{-1} \sim W\left(\underline{S}^{-1}, \underline{\nu}\right). \tag{4.8}$$

Note that, as Koop & Korobilis (2009) point out, the variance of α does not depend on Σ and its chosen values are up to the researcher. Even though the full posterior distribution does not have an analytical form, conditional distributions have form of

$$\alpha|y, \Sigma^{-1} \sim N\left(\overline{\alpha}, \overline{V}_{\alpha}\right), \tag{4.9}$$

$$\Sigma^{-1}|y, \alpha \sim W\left(\overline{S}^{-1}, \overline{\nu}\right), \tag{4.10}$$

where

$$\overline{V}_{\beta} = \left(\underline{V}_{\beta}^{-1} + \sum_{t=1}^{T} Z_t' \Sigma^{-1} Z_t\right)^{-1}, \tag{4.11}$$

$$\overline{\alpha} = \overline{V}_{\beta}\left(\underline{V}_{\beta}^{-1}\underline{\alpha} + \sum_{t=1}^{T} Z_t' \Sigma^{-1} y\right), \tag{4.12}$$

$$\overline{\nu} = \underline{\nu} + T, \tag{4.13}$$

$$\overline{S} = \underline{S} + \sum_{t=1}^{T} \left(y_t - Z_t\alpha\right)\left(y_t - Z_t\alpha\right)', \tag{4.14}$$

which allows usage of the Gibbs sampler in the following way:

1. Initialize $\underline{\alpha}, \underline{V}_{\alpha}, \underline{S}^{-1}$ and $\underline{\nu}$
2. Draw α from $p\left(\alpha|y, \Sigma^{-1}\right)$
3. Draw Σ from $p\left(\Sigma^{-1}|y, \alpha\right)$
4. Go back to 2.

4.2.2 Independent Minnesota-Wishart prior

The Independent Minnesota-Wishart prior combines the useful properties of the Minnesota prior (see below) with drawing the variance matrix Σ from the Inverted Wishart distribution.

The Minnesota prior has been created by researchers of Federal Reserve Bank of Minneapolis in the 1980's (see Litterman (1986)). Its properties made the estimation of Bayesian models much easier. The advantage of Minnesota prior lies in replacing the variance matrix Σ with a given estimate $\hat{\Sigma}$. It follows that the Bayesian inference of α does not depend on Σ but only on its OLS estimator. The posterior of α follows

$$\alpha \sim N\left(\underline{\alpha}_{Mn}, \underline{V}_{Mn}\right). \tag{4.15}$$

Following Koop & Korobilis (2010), we will set all $\underline{\alpha}_{Mn}$ to 0 as our data were differenced. Similarly, we decided to follow Koop & Korobilis (2010) in their approach to set $\underline{V}_{Mn}$. As Minnesota prior assumes $\underline{V}_{Mn}$ to be diagonal, let us denote its block corresponding to coefficients in i-th equation as $\underline{V}_i$. Moreover, $\underline{V}_{i,jj}$ denotes diagonal elements of $\underline{V}_i$. The prior for $\underline{V}_{i,jj}$ follows

$$\underline{V}_{i,jj} = \begin{cases} \frac{\underline{a}_1}{r^2} \text{ for coefficients on own lag } r \text{ for } r = 1, \dots, p \\ \frac{\underline{a}_2 \sigma_{ii}}{r^2 \sigma jj} \text{ for coefficients on lag } r \text{ of variable } j \neq i \text{ for } r = i, \dots, p \\ \underline{a}_3 \sigma_{ii} \text{ for coefficients on exogenous variables.} \end{cases} \tag{4.16}$$

This specification narrows down the complicated specification to choosing the level of scalars $\underline{a}_1$, $\underline{a}_2$ and $\underline{a}_3$. In our estimation we use the same values as Koop & Korobilis (2009), the robustness of results to their specification is presented in section 6.1.1.

The posterior of α follows the distribution

$$\alpha | y \sim N\left(\overline{\alpha}_{Mn}, \overline{V}_{Mn}\right), \tag{4.17}$$

where

$$\overline{V}_{Mn} = \left[\underline{V}_{Mn}^{-1} + \left(\hat{\Sigma}^{-1} \otimes (X'X)\right)\right]^{-1}, \tag{4.18}$$

$$\overline{\alpha}_{Mn} = \overline{V}_{Mn} \left[\underline{V}_{Mn}^{-1} \underline{\alpha}_{Mn} + \left(\hat{\Sigma}^{-1} \otimes X\right)' y\right]. \tag{4.19}$$

Similarly as in case of the INW prior, draws of Σ are obtained fom the distribution $\Sigma^{-1}|y, \alpha \sim W\left(\overline{S}^{-1}, \overline{\nu}\right)$. An analogical Gibbs sampler follows.

4.3 TVP-VAR with stochastic volatility

The conducted empirical research has proven that relaxing of the residual homoscedasticity assumption can considerably improve the model as volatility of financial time series' tends to cluster. It is not unlikely that similar statement holds as well for allowing time variation in regression coefficients. This section presents the methodology of heteroscedastic Time Varying Parameter VAR model (TVP-VAR) presented by equation

$$y_t = c_t + A_t^1 y_{t-1} + A_t^2 y_{t-2} + A_t^3 y_{t-3} + A_t^4 y_{t-4} + A_t^5 y_{t-5} + e_t, \tag{4.20}$$

where y_t and e_t have the same properties as in (4.1).[5] The crucial difference is that the constant c and matrices A^i can now be different for each time unit.

Under the frequentist approach it would be impossible to estimate such a model because of overparametrization, however, Bayesian methods deal with this issue by introducing shrinkage of coefficients (Koop & Korobilis 2010). In practice, some or all parameters are shrunk towards zero using the prior definition and then updated recursively in an MCMC algorithm.

Since the pioneering work of Canova (1993), numerous applications of TVP-VAR modelling have been performed. Estimation of our model follows the work of Primiceri (2005). As Bayesian econometrics allows for an extreme variety of possible specifications, inference and analysis, it is beyond the scope of this thesis to provide overview of specific methods that have been used until now. Section 2.4 leads an interested reader to several empirical papers that perform TVP-VAR models.

4.3.1 Rearrangement of variables

Let us assume that the residual vector e_t has the variance covariance matrix Ω_t. In order to make the estimation more efficient, this matrix Ω_t will be decomposed by a triangular reduction

$$\Lambda_t \Omega_t \Lambda_t' = \Sigma_t \Sigma_t', \tag{4.21}$$

[5] Note that, rigorously speaking, e_t has the same properties as ε_t.

where

$$\Lambda_t = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \lambda_{21,t} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \lambda_{n1,t} & \cdots & \lambda_{n(n-1),t} & 1 \end{bmatrix} \tag{4.22}$$

and Σ_t is a diagonal equation satisfying

$$diag\left(\Sigma_t\right) = \left(\sigma_{1,t}, \sigma_{2,t}, \cdots, \sigma_{n,t}\right). \tag{4.23}$$

Equation 4.20 can then be rewritten as

$$y_t = X_t' A_t + \Lambda_t^{-1} \Sigma_t \varepsilon_t, \tag{4.24}$$

where the right hand side variables are stacked into the form

$$X_t' = I_n \otimes \begin{bmatrix} 1 & y_{t-1} & \cdots & y_{t-5} \end{bmatrix}, \tag{4.25}$$

$$A_t = \begin{bmatrix} c & A_t^1 & \cdots & A_t^5 \end{bmatrix}. \tag{4.26}$$

It can be shown that under this notation $var\left(\varepsilon_t\right) = I_n$, which is a desired property.

4.3.2 Model dynamics

Selected variables of the model are estimated as random walks. Given the definition of A_t from (4.26) and creating a vector λ_t by stacking the non-zero and non-one elements of Λ_t by rows,[6] model dynamics follows equations

$$A_t = A_{t-1} + \nu_t, \tag{4.27}$$

$$\lambda_t = \lambda_{t-1} + \zeta_t, \tag{4.28}$$

$$\log \sigma_t = \log \sigma_{t-1} + \eta_t. \tag{4.29}$$

[6] Thus, $\lambda_t = \begin{bmatrix} \lambda_{21,t} & \lambda_{31,t} & \lambda_{32,t} & \lambda_{41,t} & \cdots & \lambda_{n(n-1),t} \end{bmatrix}'$.

Components of the variance matrix V take the form

$$V = \begin{bmatrix} \varepsilon_t \\ \nu_t \\ \zeta_t \\ \eta_t \end{bmatrix} = \begin{bmatrix} I_n & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & S & 0 \\ 0 & 0 & 0 & W \end{bmatrix}, \tag{4.30}$$

where Q, S and W are positive definite matrices and S is assumed to be block diagonal in blocks corresponding to coefficients of each equation.

4.3.3 Priors

Priors using OLS estimates from the training sample of 40 initial observations are set as

$$A_0 \sim N\left(\hat{A}_{OLS}, 4 \cdot V\left(\hat{A}_{OLS}\right)\right), \tag{4.31}$$

$$\Lambda_0 \sim N\left(\hat{\Lambda}_{OLS}, 4 \cdot V\left(\hat{\Lambda}_{OLS}\right)\right), \tag{4.32}$$

$$\log \sigma_0 \sim N\left(\log \hat{\sigma}_{OLS}, 4 \cdot I_n\right), \tag{4.33}$$

$$Q \sim IW\left(k_Q^2 \cdot 40 \cdot V\left(\hat{A}_{OLS}\right), 40\right), \tag{4.34}$$

$$W \sim IW\left(k_W^2 \cdot 5 \cdot I_n, 5\right), \tag{4.35}$$

$$S_1 \sim IW\left(k_S^2 \cdot 2 \cdot V\left(\hat{\Lambda}_{1,OLS}\right), 2\right), \tag{4.36}$$

$$S_2 \sim IW\left(k_S^2 \cdot 3 \cdot V\left(\hat{\Lambda}_{2,OLS}\right), 3\right), \tag{4.37}$$

$$S_3 \sim IW\left(k_S^2 \cdot 4 \cdot V\left(\hat{\Lambda}_{3,OLS}\right), 4\right), \tag{4.38}$$

where values of particular coefficients are calculated by the function *tsprior()* obtained from Koop & Korobilis (2010). Values of multiplication parameters k_Q^2, k_W^2 and k_S^2 are discussed in section 5.2.

4.3.4 Gibbs sampling

The Gibbs sampler takes draws from conditional distributions in the following way:

1. Initialize of $\Lambda^T, \Sigma^T, s^T \text{and} V$.

2. Draw A^T from $p(A^T|y^T, \Lambda^T, \Sigma^T, V)$.

3. Draw Λ^T from $p(\Lambda^T|y^T, A^T, \Sigma^T, V)$.

4. Draw Σ^T from $p(\Sigma^T|y^T, A^T, \Lambda^T s^T, V)$.

5. Draw S^T from $p(S^T|y^T, \Lambda^T \Sigma^T, V)$.

6. Draw V by drawing $Q, W \text{and} S$ from $p(Q, W, S|y^T, A^T, \Lambda^T, \Sigma^T) = p(Q|y^T, A^T, \Lambda^T, \Sigma^T) \cdot p(W|y^T, A^T, \Lambda^T, \Sigma^T) \cdot p(S_1|y^T, A^T, \Lambda^T, \Sigma^T) \cdot \ldots \cdot p(S_{n-1}|y^T, A^T, \Lambda^T, \Sigma^T)$

7. Go back to the second step and repeat.

In step 2, the algorithm of Carter & Kohn (1994) is used.

4.4 Impulse responses, variance decomposition

This section explains the methodology of obtaining impulse responses and variance decomposition in our models. Without loss of generality, all indices shown in this section apply for the basic BVAR model. Extension to the framework of TVP-VAR model simply lies in having a separate measure for each time unit in the sample.

4.4.1 Impulse responses

Impulse responses can be calculated due to the well known fact that every autoregressive function can be transformed to a moving average form. This fact holds for vector models as well, the resulting models are called Vector Moving Average (VMA) models.[7] An intuitive illustration in case of bivariate VAR is shown in Diebold & Yilmaz (2009). Let us rewrite equation 4.1 as

$$y_t = c + \sum_{i=1}^{5} A^i y_{t-i} + \varepsilon_t. \tag{4.39}$$

This can be rearranged into

$$y_t = \mu + \sum_{i=1}^{t} \Phi^i \varepsilon_{t-i}. \tag{4.40}$$

Coefficients Φ^1 to Φ^t can then be used to calculate impulse responses to the shock to the dependent variable at time t.[8] As far as the system does not follow

[7] For details on VMA models see e.g. Tsay (2002).

[8] Steps of this calculation are very technical and beyond the scope of this thesis. Hamilton (1994) provides a very detailed explanation.

an exploding path, particular coefficients will shrink towards zero with longer horizons which is in line with economic theory. Our analysis sets the horizon to 21 lags, where the first lag describes the impact at time t and the rest of coefficients represent responses of the system in four weeks following the shock.

Note that due to the step-wise calculation of impulse responses, a shock into the first variable at time t will not affect the other variables until time $t+1$, but a shock to the i-th variable will affect variables 1 to $i-1$ already at time t. Therefore, the order of variables matters in calculation of VAR impulse responses. This may potentially create issues in case of stock data, as there usually is no information about the possible direction of spillovers from one market to the other ones. The only exception for stock market data that we can think of are studies that examine data from various time zones - in such case, researcher should order the data according to the geographical location from east to west.[9] The robustness of our variable ordering is discussed in section 6.1.2.

4.4.2 Variance decomposition

Let us now define the impulse response storage matrix Φ which takes form

$$\Phi^h = \begin{bmatrix} \Phi^h_{1,1} & \Phi^h_{1,2} & \Phi^h_{1,3} & \Phi^h_{1,4} \\ \Phi^h_{2,1} & \Phi^h_{2,2} & \Phi^h_{2,3} & \Phi^h_{2,4} \\ \Phi^h_{3,1} & \Phi^h_{3,2} & \Phi^h_{3,3} & \Phi^h_{3,4} \\ \Phi^h_{4,1} & \Phi^h_{4,2} & \Phi^h_{4,3} & \Phi^h_{4,4} \end{bmatrix}, \tag{4.41}$$

where $\Phi^h_{i,j}$ marks the impulse response of the j-th variable to the shock in the i-th variable at $h-th$ period after the shock. The matrix of relative variance

[9] In macroeconomic modelling it is often the case that researcher can impose such assumptions.

decomposition Ξ takes form

$$\Xi = \begin{bmatrix} \frac{\sum_{h=1}^{21} \Phi_{1,1}^h}{\sum_{i=1}^{4}\sum_{h=1}^{21} \Phi_{i,1}^h} & \frac{\sum_{h=1}^{21} \Phi_{1,2}^h}{\sum_{i=1}^{4}\sum_{h=1}^{21} \Phi_{i,2}^h} & \frac{\sum_{h=1}^{21} \Phi_{1,3}^h}{\sum_{i=1}^{4}\sum_{h=1}^{21} \Phi_{i,3}^h} & \frac{\sum_{h=1}^{21} \Phi_{1,4}^h}{\sum_{i=1}^{4}\sum_{h=1}^{21} \Phi_{i,4}^h} \\ \frac{\sum_{h=1}^{21} \Phi_{2,1}^h}{\sum_{i=1}^{4}\sum_{h=1}^{21} \Phi_{i,1}^h} & \frac{\sum_{h=1}^{21} \Phi_{2,2}^h}{\sum_{i=1}^{4}\sum_{h=1}^{21} \Phi_{i,2}^h} & \frac{\sum_{h=1}^{21} \Phi_{2,3}^h}{\sum_{i=1}^{4}\sum_{h=1}^{21} \Phi_{i,3}^h} & \frac{\sum_{h=1}^{21} \Phi_{2,4}^h}{\sum_{i=1}^{4}\sum_{h=1}^{21} \Phi_{i,4}^h} \\ \frac{\sum_{h=1}^{21} \Phi_{3,1}^h}{\sum_{i=1}^{4}\sum_{h=1}^{21} \Phi_{i,1}^h} & \frac{\sum_{h=1}^{21} \Phi_{3,2}^h}{\sum_{i=1}^{4}\sum_{h=1}^{21} \Phi_{i,2}^h} & \frac{\sum_{h=1}^{21} \Phi_{3,3}^h}{\sum_{i=1}^{4}\sum_{h=1}^{21} \Phi_{i,3}^h} & \frac{\sum_{h=1}^{21} \Phi_{3,4}^h}{\sum_{i=1}^{4}\sum_{h=1}^{21} \Phi_{i,4}^h} \\ \frac{\sum_{h=1}^{21} \Phi_{4,1}^h}{\sum_{i=1}^{4}\sum_{h=1}^{21} \Phi_{i,1}^h} & \frac{\sum_{h=1}^{21} \Phi_{4,2}^h}{\sum_{i=1}^{4}\sum_{h=1}^{21} \Phi_{i,2}^h} & \frac{\sum_{h=1}^{21} \Phi_{4,3}^h}{\sum_{i=1}^{4}\sum_{h=1}^{21} \Phi_{i,3}^h} & \frac{\sum_{h=1}^{21} \Phi_{4,4}^h}{\sum_{i=1}^{4}\sum_{h=1}^{21} \Phi_{i,4}^h} \end{bmatrix}, \tag{4.42}$$

where a term in i-th row and j-th column denotes the relative importance of shocks coming from i-th variable into j-th variable. Note that terms in each column will always sum to unity.

4.4.3 Spillover indices

The results of variance decomposition can be used to calculate the Spillover Index (SI) defined by Diebold & Yilmaz (2009). If we denote the term in i-th row and j-th collumn of the relative variance decomposition matrix Ξ as $\Xi_{i,j}$, SI takes the form

$$S = \frac{\sum_{i=1}^{4} \sum_{j=1, i\neq j}^{4} \Xi_{i,j}}{\sum_{i=1}^{4} \sum_{j=1}^{4} \Xi_{i,j}} \tag{4.43}$$

The Spillover Index measures how much of the total variance in the sample is caused by spillovers among markets to another. Following the results of Diebold & Yilmaz (2009), it is expected to be higher in times of crises and lower in stable times.

4.4.4 Impulse performance diagnostics

In order to compare performance of impulse responses of several models, we use the following measure. Let us define the *Impulse Performance Diagnostics* (IPD) for each country as

$$IPD_{ctr} = \sum\nolimits_{i=1}^{M} \sum\nolimits_{j=1}^{h} \left(q(\Phi_{ctr,i}^{j}, q_u) - q(\Phi_{ctr,i}^{j}, q_b) \right)^2, \tag{4.44}$$

where $\Phi^{j}_{ctr,i}$ is an impulse response defined by (4.41), q_u and q_b are the specified upper and bottom quantiles and h is the horizon of IPD calculation. The resulting diagnostics is a $1\times(M+1)$ vector $IPD = \begin{bmatrix} IPD_1 & \cdots & IPD_M & IPD_{tot} \end{bmatrix}$ where the term IPD_{tot} is the sum of M previous rows.

As the diagnostics measures the square distance between two specified quantiles of impulse response distribution, the model with the lowest value of IPD can be seen as the best model in terms of impulse responses.

4.5 Matlab programming

As our models require usage of numerical software, Matlab codes were downloaded from the website associated with Koop & Korobilis (2009).[10] These codes were adjusted in order to comply with our notation and methodology. Apart from code adjustment, several additional functionalities were programmed in order to process the estimation results. This section describes the two most important programmes that were coded. Both of them concern manipulation with draws of impulse response functions. Rest of the created codes including instructions how to use them are available by request.

4.5.1 Extracting impulse responses

Once the necessary quantiles of impulse responses have been calculated, one does not need to store full draws of impulse response functions. Therefore, we programmed a code which extracts impulse response draws from the specified files and saves these results into a new file. This considerably helps with processing of the data, as it reduces the needed computer memory by a factor of "number of replications / number of quantiles".

4.5.2 Impulse responses for each time period

The code of Koop & Korobilis (2009) calculates impulse responses for three specific periods in the TVP-VAR model. As a generalization, we programmed the code to calculate them for each time period of the sample. As such calculation needs to store all impulse response draws for each time unit of the sample, enormous computer memory is required. For example, in case of 50000 Gibbs sampler iterations, approximately 80 GB of memory is needed, which is around

[10] http://personal.strath.ac.uk/gary.koop/bayes_matlab_code_by_koop_and_korobilis.html

40 times more than a 32-bit MS Windows can handle. This issue was solved by saving sequential stores into files of a size that can be specified by the user. All draws are then combined together with sequential opening and closing of all saved files at each time unit.

Chapter 5

Empirical Results

5.1 BVAR

The Independent Normal-Wishart and Independent Minnesota-Wishart[1] prior models were estimated with 2000 burn-in draws and 10000 normal draws. The convergence was confirmed as their results are almost identical to the model with 20000 + 50000 iterations (see section 6.1.4). Values of parameters $\underline{a}_1$, $\underline{a}_2$ and $\underline{a}_3$ in the IMW prior were set according to the discussion in section 6.1.1.

5.1.1 Impulse responses

Impulse responses for all four variables are depicted in figures 5.1 and 5.2. The INW benchmark model was selected according to the best performance from tests for four different variable orderings (see section 6.1.2). The benchmark IMW model was chosen with the same parameters $\underline{a}_1$, $\underline{a}_2$ and $\underline{a}_3$ as in Koop & Korobilis (2009) (For robustness, see section 6.1.1).

Note that even though impulse responses obtained from both priors produce very similar results, we can see that the 10-th and 90-th confidence bands are tighter in case of INW prior. Table 5.1 confirms this rigorously, as the IPD diagnostics for the two benchmark models clearly suggests that the INW model outperforms the second model in terms of impulse responses.

We can see that responses converge to zero in the long term, which is in line with economic theory of diminishing effects of shocks. Specifically, all shocks to Central European markets tend to disappear in the horizon of approximately one and half weeks after the shock. Moreover, all markets tend to slightly

[1]From now on, we will use the abbreviation INW for the Independent Normal-Wishart prior. Abbreviation IMW applies analogically.

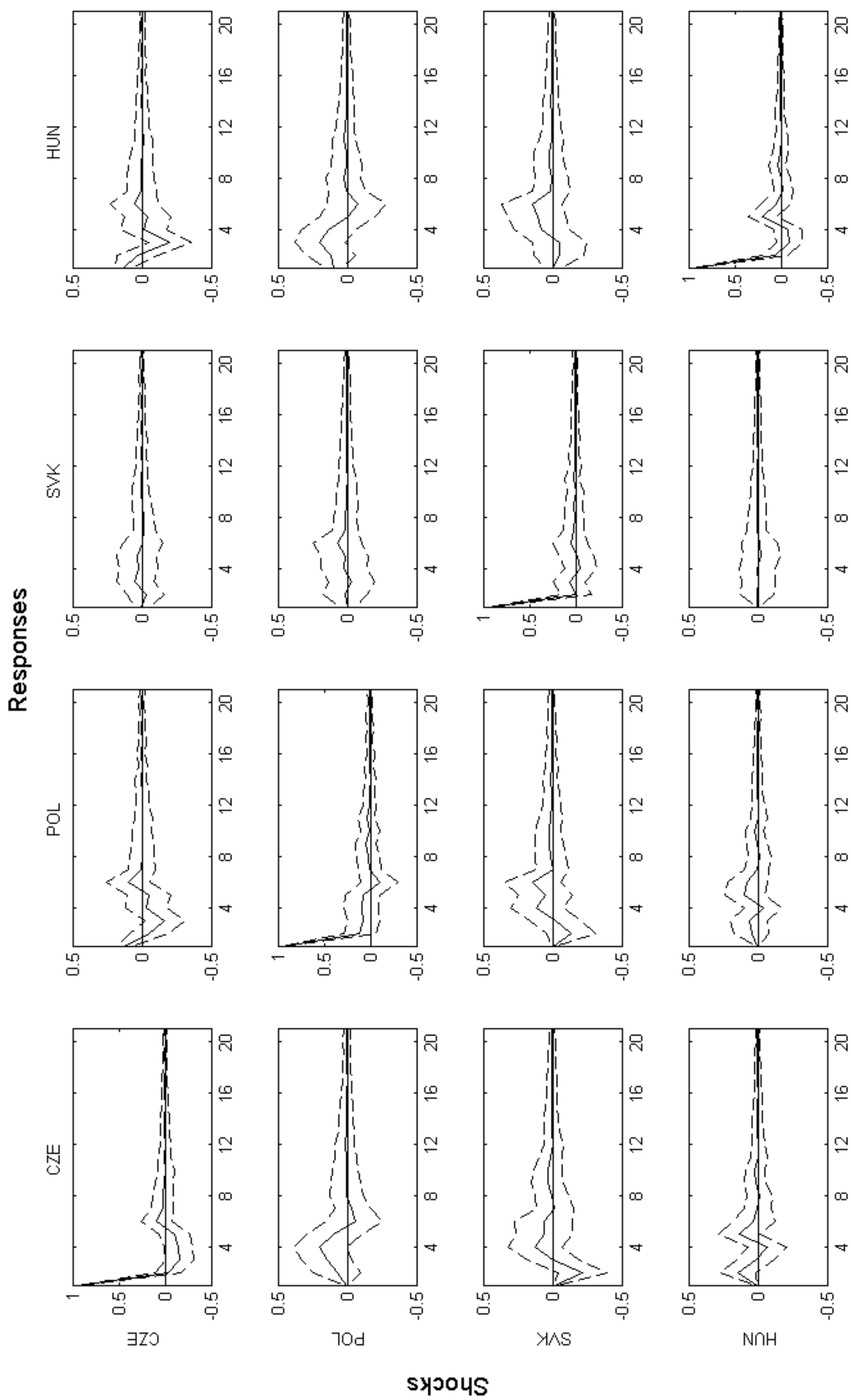

Figure 5.1: Impulse responses, Independent Normal-Wishart prior

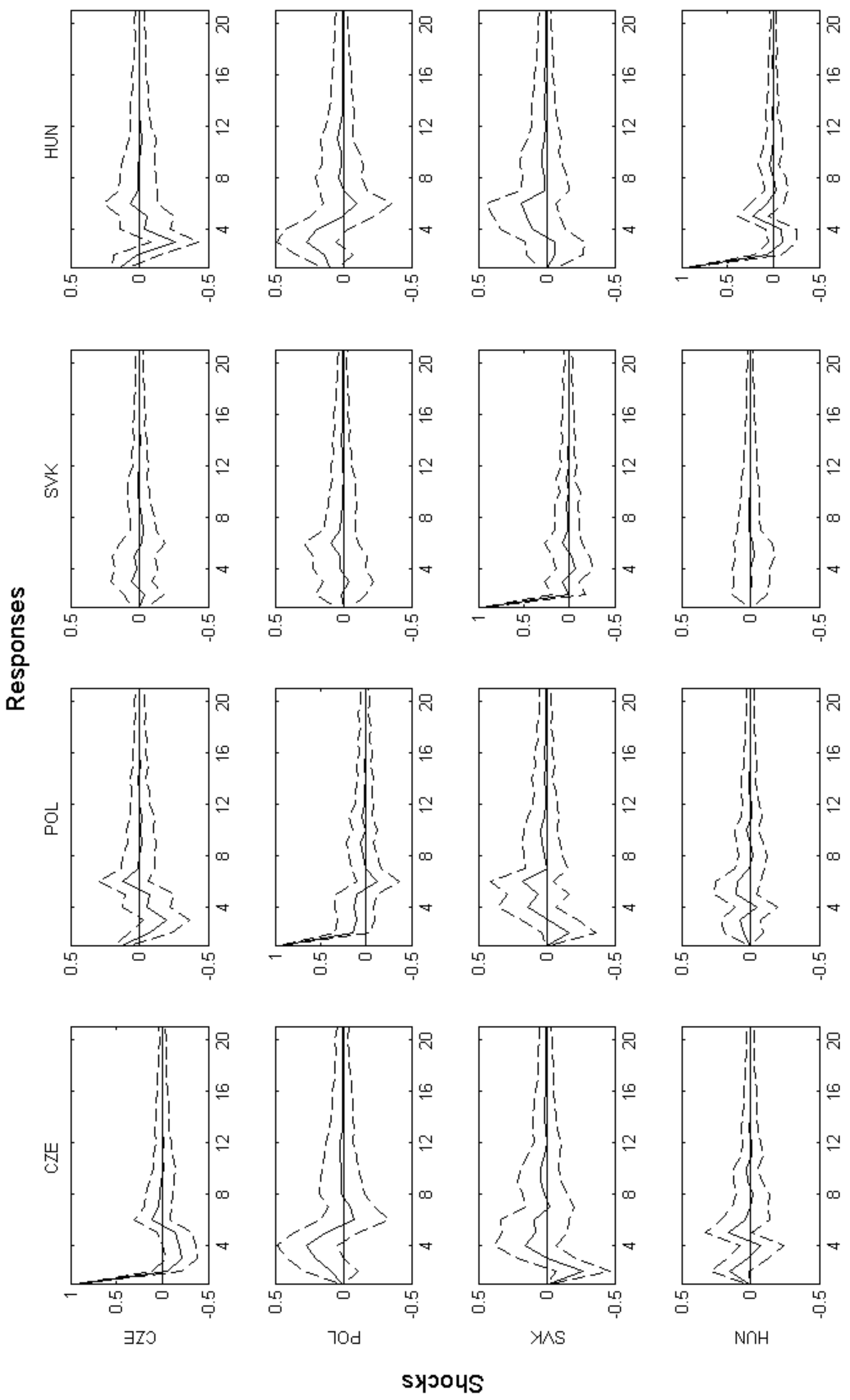

Figure 5.2: Impulse responses, Independent Minnesota-Wishart prior

Table 5.1: IPD diagnostics of benchmark models

	Shock to				
Prior	CZE	POL	SVK	HUN	Total shock
INW	3.4064	3.156	2.5569	3.5927	12.7120
IMW	5.2570	4.6168	3.4972	5.4526	18.8236

Note: IPD diagnostics measures the square distance between 10-th and 90-th quantiles of impulse response distribution.

overreact to shocks in any other market in a period of one week or sooner. Similarly as in section 4.1, we can see that the behavior of Slovak market is different from the other markets as it generally does not seem to be responding to shocks in either of the other countries. The fact that, from the statistical point of view, Slovak market moves almost randomly, can be explained by relatively low liquidity on this market.

The strange behavior of Slovak market also results to the finding that positive shocks in Slovakia will cause all three other markets to fall. From the portfolio diversification point of view, it might be interesting to invest into the Slovak market as it is almost independent of the other Central European countries. An interesting finding is that Polish and Hungarian markets react negatively to shocks in the Czech republic as well, which can be a sign that investors view these three markets separately.

To the contrary, all markets tend to react positively to shocks in Poland and Hungary, suggesting that investors do not see an opportunity for portfolio diversification among Czech, Polish and Hungarian markets. To summarize, our impulse response analysis finds out that Slovak market stands separately from the other three markets, but the evidence for behavior of the other three markets is ambiguous.

5.1.2 Variance decomposition

Figure 5.3 depicts the relative variance decomposition of model with INW prior. Czech Republic is depicted by the blue color, Poland is red, Slovak is yellow and Hungary is green (These colors will stay the same for the remainder of this thesis). As the differences in impulse response analysis are not crucial, we do not report results of variance decomposition from IMW prior in this chapter. These are reported in section 6.1.5.

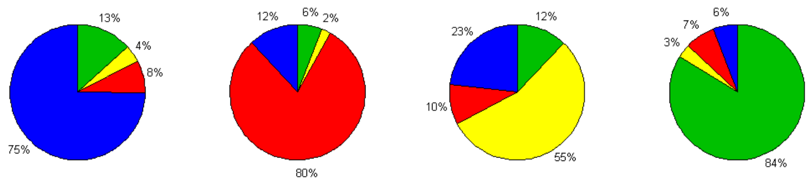

Figure 5.3: Relative variance decomposition, INW prior

Table 5.2 shows the variance decomposition of the benchmark model. Results show an expected fact that the variance in each country is driven mainly on its own. This also holds for Slovakia with the lowest estimated share of 55%. We find possible explanation for Slovakia's lowest share twofold: Firstly, it is the only country that shares its borders with all the other three countries. This should not be a reason under the assumption of zero transaction costs, however, even though such costs have become lower with the recent technological boom, they still undoubtedly exist. Secondly, market capitalisation of Bratislava Stock Exchange is by far the lowest out of the four countries in our sample, reaching the level of 29% of the third Hungary and only 2.69% of the total market capitalisation in sample countries in 2011.[2]

The Spillover Index of 26.43% means that approximately one quarter of total volatility in Central European markets was caused by spillovers among them. Interestingly, Polish stock exchange with the biggest market capitalization did not report either the highest contribution to others or the lowest contribution from others. The biggest contribution to others (and in total) comes from the Czech Republic and the lowest contribution from others is reported from Hungary. This can be explained by the geographical reason of Hungary only neighbouring Slovakia, the country with the lowest market capitalisation and liquidity. It is also possible that such differences were caused by different policy responses in times of crisis.

[2]Data were obtained from Eurostat database in April 2012.

Table 5.2: Variance decomposition

Market	CZE	POL	SVK	HUN	Contribution from others
CZE	74.99	7.64	4.02	13.35	25.01
POL	11.87	80.15	2.23	5.75	19.85
SVK	23.12	9.53	55.3	12.05	44.7
HUN	5.76	7.06	3.35	83.82	16.18
Contribution to others	40.75	24.23	9.6	31.15	
Total contribution	115.74	104.38	64.9	114.98	

Note: 21 day horizon. All numbers in percentages. Spillover index: 26.43%

Previous results reported the relative variance decomposition in order to show which share of volatility is caused by which market. Figure 5.4 extends these data by including levels of total variance in particular countries. Interestingly, Slovakia reports the highest volatility and at the same time lowest contributions to the other countries. Once again, this can be explained by the relative isolation from the other markets and by its low liquidity. Nevertheless, we can see that the relationship of Slovakia to Czech Republic is stronger than to Hungary, which can be explained by their historical proximity.

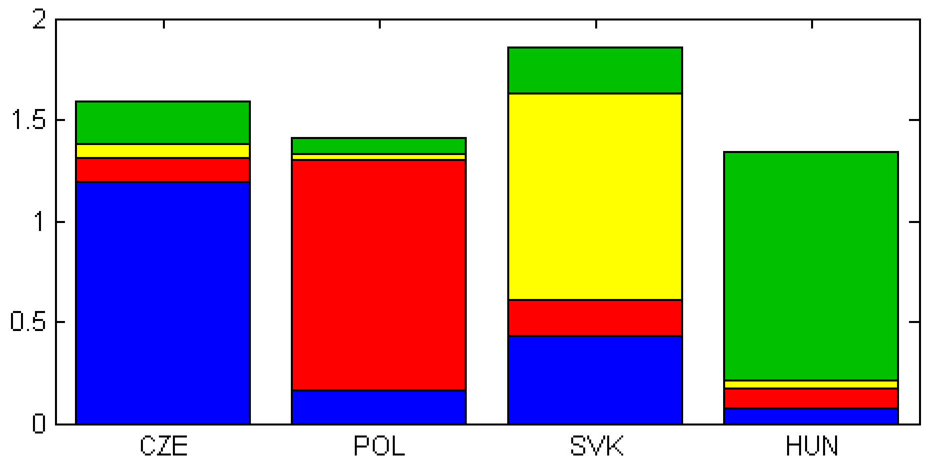

Figure 5.4: Absolute variance decomposition, INW prior

Note that we do not report models for the three subsamples defined in chapter 4. The reason for this is purely practical, as the estimation was not able to produce converging impulse responses even for 100000 burn-in and 10000 iterations.

5.2 TVP-VAR with stochastic volatility

Even though we have tried many different models, convergence of impulse responses in the time-varying model has not been achieved. However, this is not an unusual result, as responses of TVP-VAR models are often exploding (Primiceri 2005), which is indeed our case. This section shows that the TVP-VAR model can be seen as an improvement to the basic BVAR model as it shows that regression coefficients and standard deviations of residuals do vary in time. The model described in this section shows results of estimation with 10000 + 10000 iterations, however, even the model with 80000 + 50000 iterations did not produce better results (see section 6.2).

5.2.1 Prior hyperparameters

The value of hyperparameters k_Q, k_W and k_S is crucial for the performance of TVP-VAR model (Primiceri 2005). Our model sets them as $k_Q = 0.03, k_W = 0.01$ and $k_S = 5$, which is little different from Primiceri's values. The coefficient k_Q has been tripled as the matrix Q was positive indefinite with lower values. This can be explained by the fact that log differences in the daily financial data are very likely to be lower than in a quarterly macroeconomic data. Coefficient k_W remained set at the value of Primiceri (2005), as we believe that there should be no differences in heteroscedasticity of residuals in financial and macroeconomic time series. The value of parameter k_S was set to 5 after the arbitrary selection among models with values of $0.1, 5$ and 100, as results of the middle model seemed most plausible. More details can be found in section 6.2.2.

5.2.2 Regression coefficients

Figures 5.6 and 5.7 show that the estimated coefficients do vary in time, which justifies the change from BVAR to TVP-VAR. As can be seen from figures in appendix 6.2, the reported coefficients vary in time in all estimated models.

5.2.3 Stochastic volatility

Figure 5.5 shows the time behavior of standard deviations of residuals for each equation. The time changing variance justifies inclusion of the stochastic volatility concept into our model. Note that the mean of standard deviations increases in times of crisis, which is an expected result.

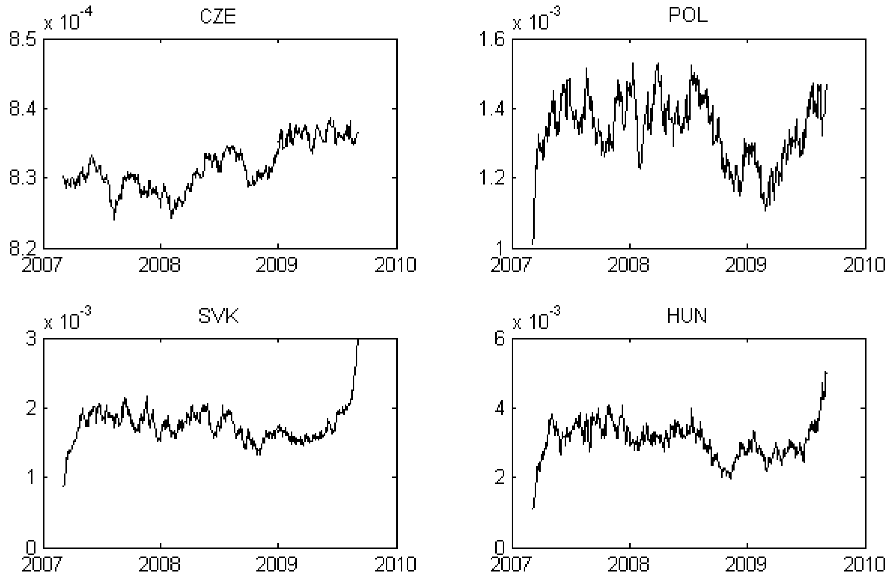

Figure 5.5: Mean of the standard deviations of residuals in time

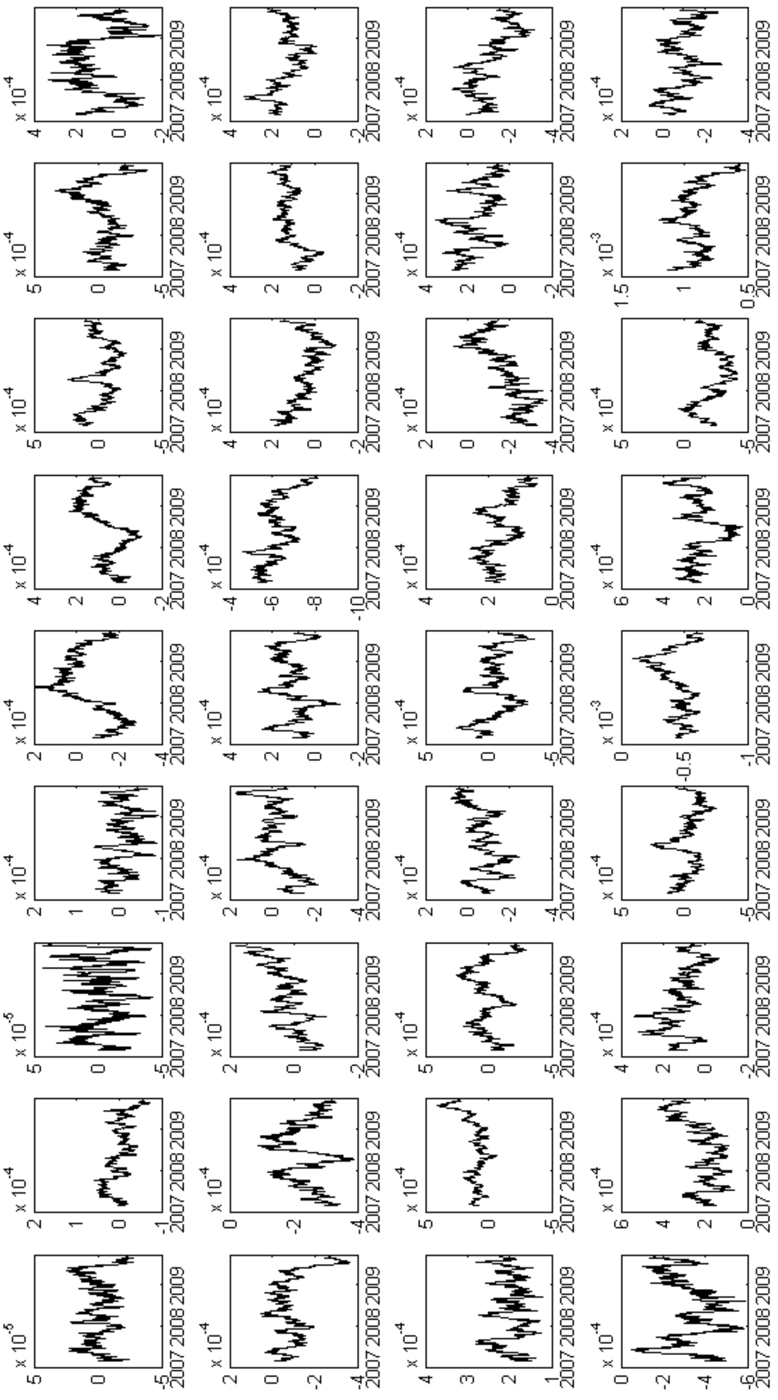

Figure 5.6: Time behavior of equation coefficients, lags 1 and 2

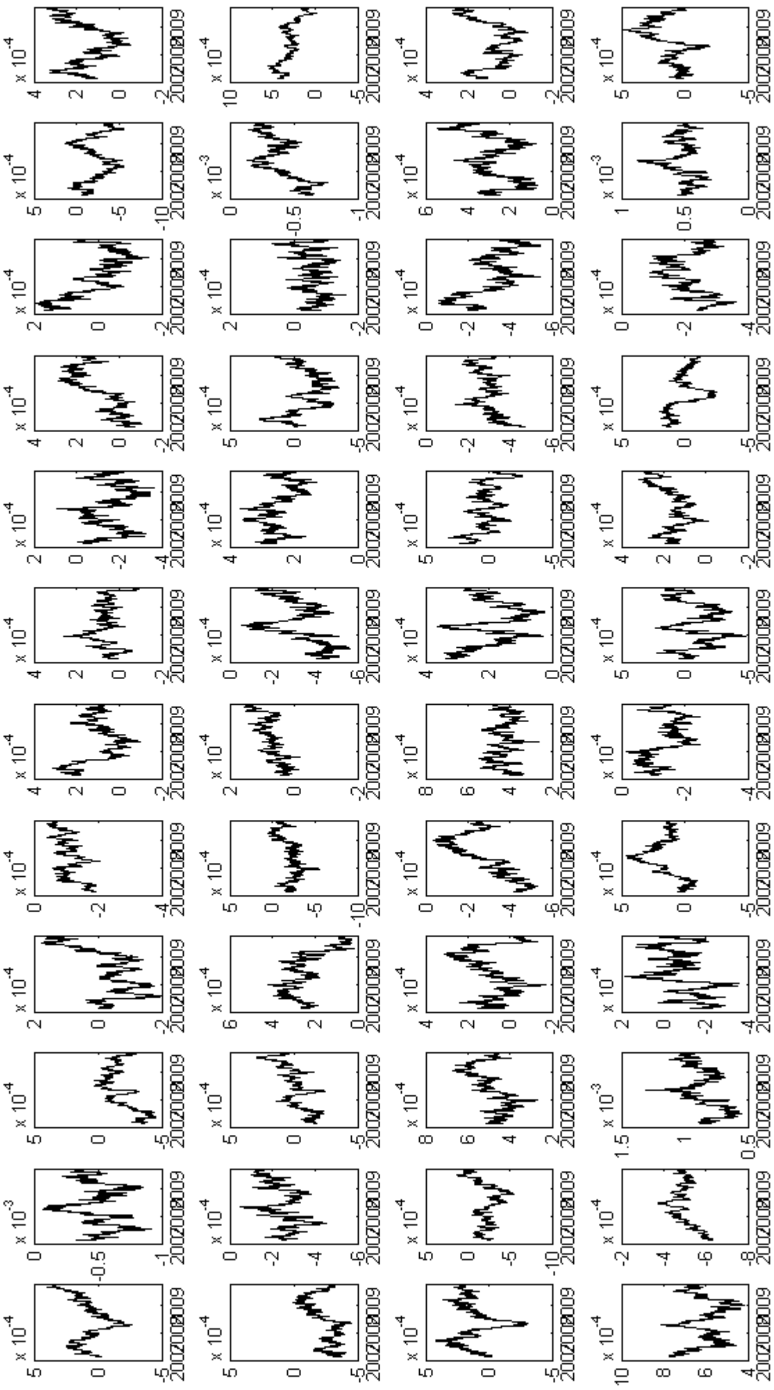

Figure 5.7: Time behavior of equation coefficients, lags 3 to 5

Chapter 6

Model Selection and Robustness

6.1 BVAR

This section shows the results of various robustness tests ran in the BVAR model. As noted in section 5.1.1, the results of INW and IMW priors are very similar, however, INW prior generally produces little better results as its confidence bands are narrower.

6.1.1 IMW parameters

As written in section 5.1, the benchmark IMW model was chosen with parameters $\underline{a}_1$, $\underline{a}_2$ and $\underline{a}_3$ as in Koop & Korobilis (2009). Table 6.1 shows results of two alternative models, each of them dividing or multiplying the parameters of the original model by a factor of 10. As a model with ten times lower parameters reported the 0.5% change of IPD diagnostics and the model with ten times higher parameters resulted in its change by 1%, we can conclude that results of IMW prior are robust to specification of its parameters.

Table 6.1: Robustness of IMW model

a1	a2	a3	CZE	POL	SVK	HUN	IPD
0.05	0.05	10	5.2462	4.6777	3.5352	5.4644	18.9234
0.5	0.5	100	5.2570	4.6168	3.4972	5.4526	18.8236
5	5	1000	5.1356	4.6635	3.4841	5.3494	18.6327

6.1.2 Order of variables

As noted in section 4.4.1, impulse responses of the same model could be different upon variable ordering. Due to the fact that, in a general VAR model with N variables, $N!$ different orders of variables are possible, it could be very computationally demanding to check for all orderings. In our case, it would require to estimate 24 different orderings for each prior.

Our approach follows Diebold & Yilmaz (2009) who perform robustness test by sequentially moving the 1-st, 2-nd to (M-1)-th variable into the last place in the sample. Results of the four models are available in table 6.2. Model in the first row was selected as a benchmark model based on the IPD diagnostics. In order to save space, we do not report impulse response figures here (see appendix A), but merely state that they also confirm the robustness for order of variables. This is also confirmed by figure 6.1, which shows the results of variance decomposition of the four models.

Table 6.2: Robustness to variable ordering

Sample	Var 1	Var 2	Var 3	Var 4	IPD	Figure
Full	CZE	POL	SVK	HUN	12.7120	5.1
Full	CZE	SVK	HUN	POL	12.7484	A.1
Full	POL	SVK	HUN	CZE	12.7536	A.2
Full	CZE	POL	HUN	SVK	12.7863	A.3

6.1.3 Horizon of IPD calculation

Table 6.3 shows the calculated IPD diagnostics for horizons from the 2-nd to the 6-th horizon. We can see that already the 6-th lag in the IMW prior reports higher IPD diagnostics than the full 20-th lag used in table 5.1 so the diagnostics is robust to the specified lag order.

Table 6.3: Robustness of IPD horizon

Prior	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6
INW	1.5995	3.2662	5.0637	6.9962	9.0540	9.8290
IMW	1.9688	4.0821	6.4447	9.0157	11.8069	13.1027

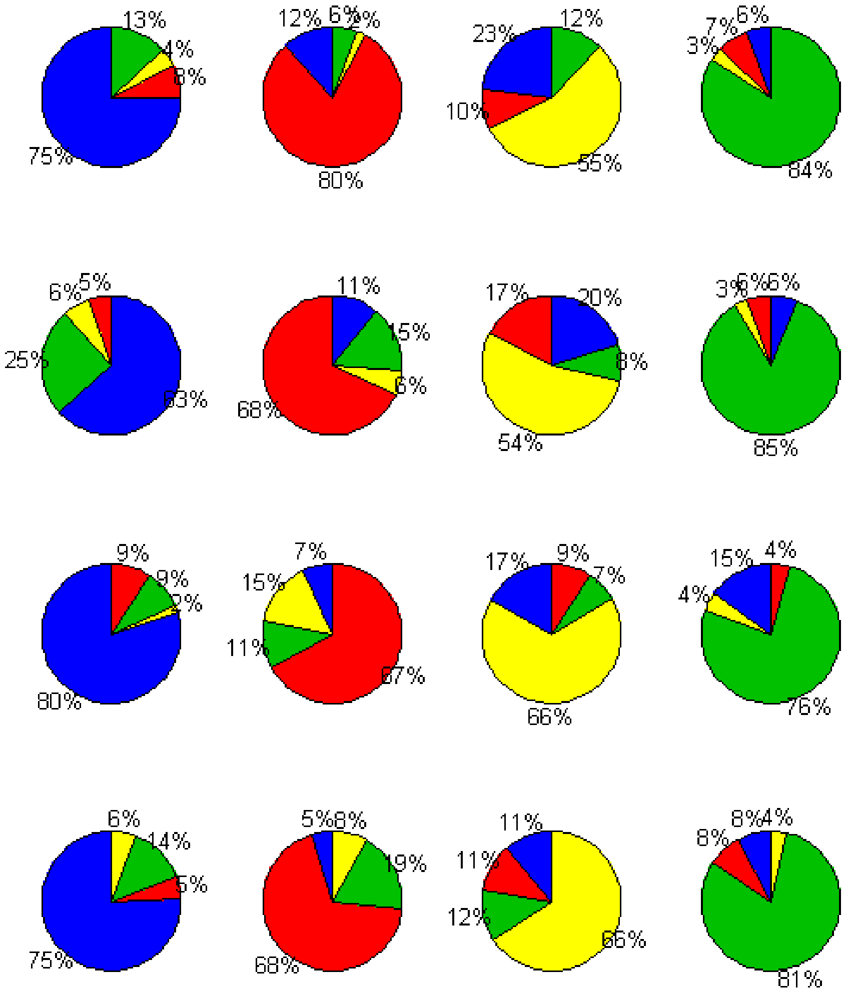

Figure 6.1: Variance decomposition, robustness to ordering

6.1.4 Number of iterations

In all applications of Gibbs sampling, the number of burn-in iterations needs to be set to a value that achieves convergence of posterior distributions. Our values of 2000 burn-in and 10000 standard iterations might seem low, however, table 6.1.4 and figure 6.2 show the IPD diagnostics and impulse responses of the INW prior with 20000 burn-in and 50000 standard iterations. As the differences between both models are marginal (0.3 % in terms of the IPD index), we can conclude that the chosen number of iterations is satisfactory.

Table 6.4: Robustness to number of iterations

Burn-in	Save	IPD_{cze}	IPD_{pol}	IPD_{svk}	IPD_{hun}	IPD_{tot}
2000	10000	3.4064	3.1560	2.5569	3.5927	12.7120
50000	20000	3.4046	3.1344	2.5548	3.5814	12.6752

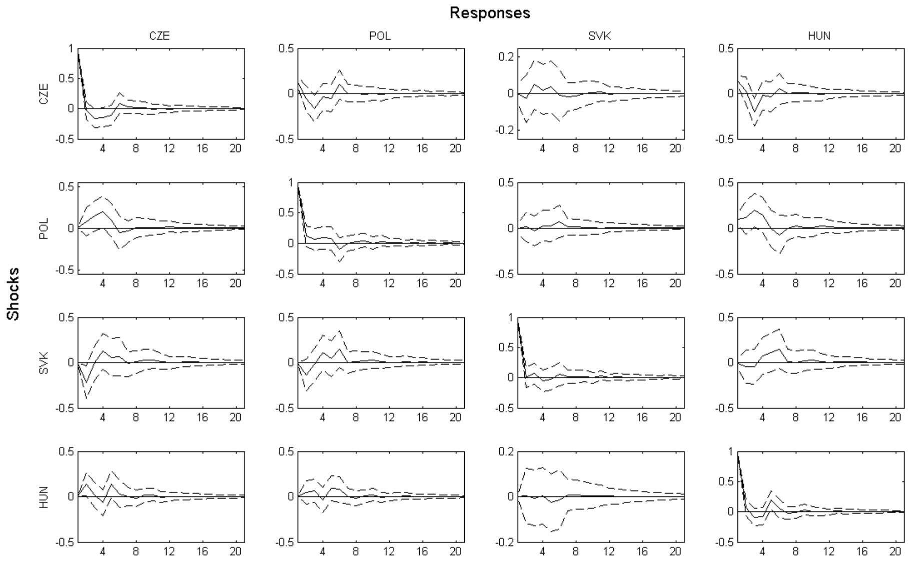

Figure 6.2: Impulse responses, INW prior, 20 000 + 50 000 iterations

6.1.5 Variance decomposition with IMW prior

Figures 6.3 and 6.4 show the results of relative and absolute variance decomposition with the chosen IMW prior. We can see that even though it produces somewhat different outcomes, essential properties of the results still stay the same. The variance of all countries is still driven mainly on its own, with Slovak market having the lowest share and at the same time the highest absolute variance.

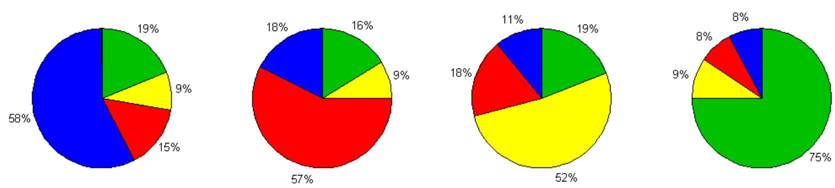

Figure 6.3: Relative variance decomposition, IMW prior

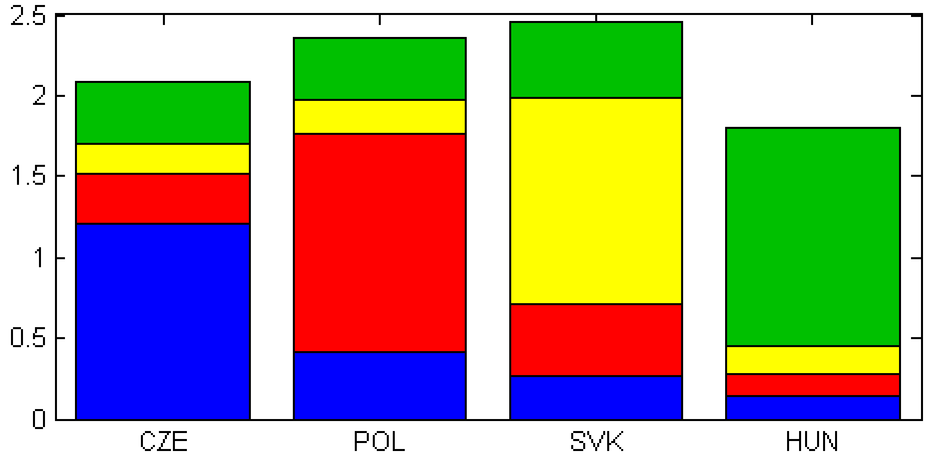

Figure 6.4: Absolute variance decomposition, IMW prior

Table 6.5: Variance decomposition, IMW prior

Market	CZE	POL	SVK	HUN	Contribution from others
CZE	57.79	14.51	8.89	18.81	42.21
POL	17.7	57.23	8.91	16.17	42.77
SVK	10.98	18.1	51.88	19.03	48.12
HUN	7.77	7.92	9.4	74.91	25.09
Contribution to others	36.45	40.53	27.19	54.01	
Total contribution	94.25	97.75	79.07	128.93	

Note: 21 day horizon. All numbers in percentages. Spillover index: 39.55%

Table 6.5 shows the numerical results of variance decomposition from IMW prior. We can see that the Slovak market still has the highest contribution from others and lowest contribution to others, Hungary keeps has lowest contribution from others. The biggest difference is that the Czech market falls from the first to the third place in both contribution to others and total contribution. However, as results from impulse response analysis reported ambiguous relationship of Czech, Polish and Hungarian markets, we do not consider this change as crucial.

6.2 TVP-VAR

As we were not able to find the specification of a TVP-VAR model that would result in non-exploding impulse responses, many different combinations of prior hyperparameters were estimated. Apart from the model described in section 5.2, this section presents results of two additional models that generally describe properties of the other estimated models. Details of these two models are shown in table 6.6.

Table 6.6: TVP robustness models

	Burn-in	Saved	k_Q	k_W	k_S	Figures
Model 1	80000	50000	0.1	0.01	0.1	6.5, A.4 and A.5
Model 2	10000	20000	10000	10000	10000	6.6, A.6 and A.7

6.2.1 Regression coefficients

Without an exception, all estimated TVP-VAR models showed distinct time variation of regression coefficients. As their figures take a lot of space, we report them in appendix A.

6.2.2 Stochastic volatility

The standard deviations of residuals from the two models described in table 6.6 are depicted in figures 6.5 and 6.6. We can see that both models differ a lot. The high value of coefficients causes the variance residuals to jump to a certain high level and then vary around such level during the whole period. The low value of residuals causes them to run smoothly in a time developing trend, however, such a smooth development is not comparable with empirical experience of financial models. As a result, a model that has coefficients inbetween of these two robust models was chosen.

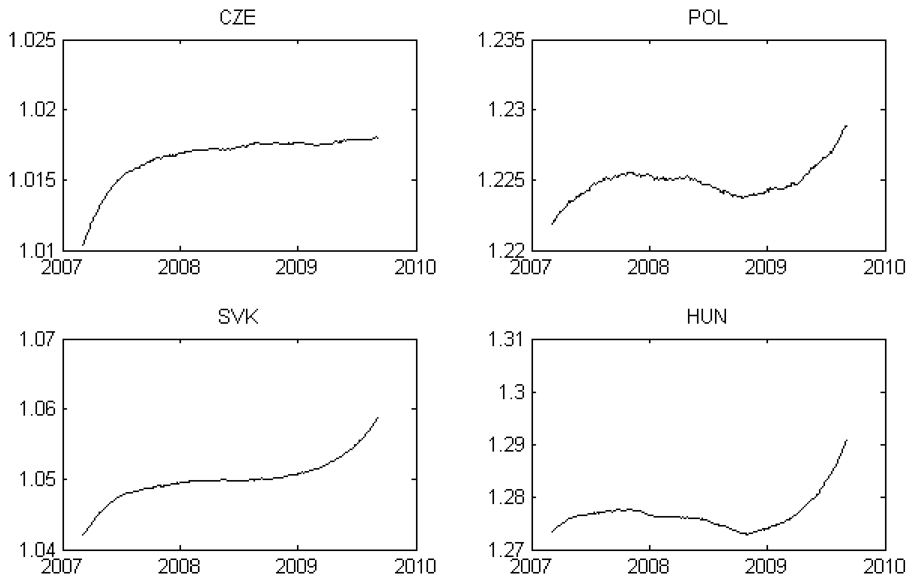

Figure 6.5: Residual time variance, robust model 1

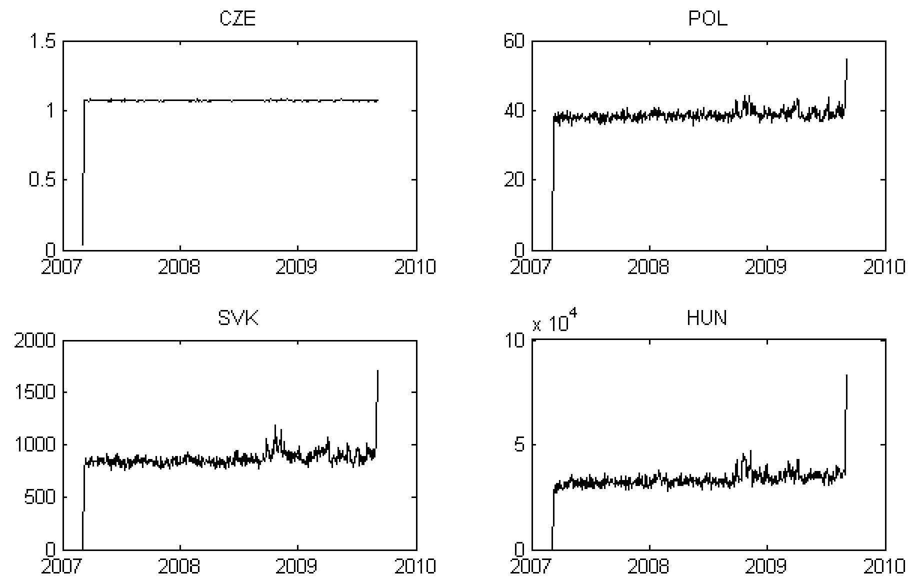

Figure 6.6: Residual time variance, robust model 2

Chapter 7

Conclusion

Impulse responses of the BVAR model show that shocks in Central European markets die out in the horizon of approximately one and half weeks. The relatively low liquidity of the Slovak market causes it to behave differently, with almost zero reactions to shocks in other countries. On the other hand, shocks in Slovak markets cause all the other markets to fall, which can be a sign that investors take the Slovak and other markets separately. The negative response of Polish and Hungarian markets to positive news in the Czech Republic strenghtens this interpretation.

The variance decomposition reveals that volatility in all markets is driven by shocks originating inside of the particular country, which is a result that is in line with economic theory. Interestingly, the Czech Republic seems to be the most influential market in terms of total volatility contribution. The Czech market is also the highest source of volatility in Slovakia, a result that possibly arises from general interconnectedness of the two economies.

Results of the TVP-VAR model show that the regression coefficients as well as volatility of residuals change in time. As expected, the estimated volatility rises in times of crisis. Unfortunately, due to the extreme computational demand of the TVP-VAR model impulse responses follow an explosive path, which is however a common result.

Posibble suggestions for future research are twofold. Firstly, next steps would be taken in order to find a way how to make the impulses of TVP model converge (on the other hand, one can argue that computational demand is the reason why, up to our knowledge, no comparable research has been yet conducted). Secondly, it would be interesting to perform the analysis with inclusion of additional countries.

Bibliography

Baumeister, C., E. Durinck, & G. Peersman (2008): "Liquidity, inflation and asset prices in a time-varying framework for the Euro area." *Working Paper Research 142*, National Bank of Belgium.

Bauwens, L., M. Lubrano, & J. Richard (1999): *Bayesian Inference in Dynamic Econometric Models*. Advanced Texts in Econometrics. Oxford University Press.

Brooks, C. (2008): *Introductory Econometrics for Finance*. Cambridge University Press.

Canova, F. (1993): "Modelling and forecasting exchange rates with a Bayesian time-varying coefficient model." *Journal of Economic Dynamics and Control* **17(1-2)**: pp. 233 – 261.

Canova, F. & M. Ciccarelli (2006): "Estimating multi-country VAR models." *Working Paper Series 603*, European Central Bank.

Canova, F. & L. Gambetti (2009): "Structural changes in the US economy: Is there a role for monetary policy?" *Journal of Economic Dynamics and Control* **33(2)**: pp. 477 – 490.

Cappiello, L., B. Gérard, A. Kadareja, & S. Manganelli (2006): "Financial integration of new EU Member States." *Working Paper Series 683*, European Central Bank.

Carter, C. K. & R. Kohn (1994): "On Gibbs sampling for state space models." *Biometrika* **81(3)**: pp. 541–553.

Chelley-Steeley, P. L. (2005): "Modeling equity market integration using smooth transition analysis: A study of Eastern European stock markets." *Journal of International Money and Finance* **24(5)**: pp. 818 – 831.

Cogley, T. & T. J. Sargent (2001): "Evolving Post-World War II U.S. Inflation Dynamics." In "NBER Macroeconomics Annual 2001, Volume 16," NBER Chapters, pp. 331–388. National Bureau of Economic Research, Inc.

Cogley, T. & T. J. Sargent (2005): "Drift and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S." *Review of Economic Dynamics* **8(2)**: pp. 262–302.

Del Negro, M. & F. Schorfheide (2004): "Priors from General Equilibrium Models for VARS." *International Economic Review* **45(2)**: pp. 643 – 673.

Diebold, F. X. & K. Yilmaz (2009): "Measuring Financial Asset Return and Volatility Spillovers, with Application to Global Equity Markets." *The Economic Journal* **119(534)**: pp. 158–171.

Dorfman, J. H. (1997): *Bayesian Economics Through Numerical Methods.* Springer.

Fama, E. F. (1970): "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance* **25(2)**: pp. 383–417.

Fama, E. F. (1998): "Market efficiency, long-term returns, and behavioral finance." *Journal of Financial Economics* **49(3)**: pp. 283 – 306.

Fedorova, E. & K. Saleem (2010): "Volatility Spillovers between Stock and Currency Markets: Evidence from Emerging Eastern Europe." *Czech Journal of Economics and Finance (Finance a uver)* **60(6)**: pp. 519–533.

Frankfurter, G. M. & E. G. McGoun (1999): "Ideology and the theory of financial economics." *Journal of Economic Behavior & Organization* **39(2)**: pp. 159–177.

Égert, B. & E. Kočenda (2011): "Time-varying synchronization of European stock markets." *Empirical Economics* **40(2)**: pp. 393–407.

Geweke, J. (1999): "Using simulation methods for Bayesian econometric models: Inference, development and communication." *Econometric Review* **18**: pp. 1–126.

Gilmore, C. G. & G. M. McManus (2002): "International portfolio diversification: US and Central European equity markets." *Emerging Markets Review* **3(1)**: pp. 69 – 83.

GRANGER, C. W. J. & P. NEWBOLD (1974): "Spurious Regressions in Econometrics." *Journal of Econometrics* **2(2)**: pp. 111–120.

GREENE, W. H. (2002): *Econometric Analysis.* Prentice Hall.

HAMILTON, J. D. (1994): *Time series analysis.* Princeton: Princeton University Press.

INGRAM, B. F. & C. H. WHITEMAN (1994): "Supplanting the OCsMinnesotaOCO prior: Forecasting macroeconomic time series using real business cycle model priors." *Journal of Monetary Economics* **34(3)**: pp. 497 – 510.

ITO, M. & A. NODA (2012): "Dynamic Stock Market Linkages and Market Efficiency." *ArXiv e-prints* .

JEANTHEAU, T. (2004): "A link between complete models with stochastic volatility and ARCH models." *Finance and Stochastics* **8(1)**: pp. 111–131.

KASCH-HAROUTOUNIAN, M. & S. PRICE (2001): "Volatility in the transition markets of Central Europe." *Applied Financial Economics* **11(1)**: pp. 93 – 105.

KIM, S., N. SHEPHARD, & S. CHIB (1998): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models." *Review of Economic Studies* **65(3)**: pp. 361–93.

KOOP, G. (2003): *Bayesian econometrics.* John Wiley & Sons.

KOOP, G. & D. KOROBILIS (2009): "Manual to accompany MATLAB package for Bayesian VAR models." *Technical report.*

KOOP, G. & D. KOROBILIS (2010): "Bayesian Multivariate Time Series Methods for Empirical Macroeconomics." *Working Paper Series 47_09*, Rimini Centre for Economic Analysis.

KOOP, G., R. LEON-GONZALEZ, & R. W. STRACHAN (2009): "On the evolution of the monetary policy transmission mechanism." *Journal of Economic Dynamics and Control* **33(4)**: pp. 997–1017.

KOOP, G., D. POIRIER, & J. TOBIAS (2007): *Bayesian econometric methods.* Econometric exercises. Cambridge University Press.

Kumar, M. (2010): "A Time-Varying Parameter Vector Autoregression Model for Forecasting Emerging Market Exchange Rates." *International Journal of Economic Sciences and Applied Research* **3(2)**: pp. 21–39.

Litterman, R. B. (1986): "Forecasting with Bayesian Vector Autoregressions-Five Years of Experience." *Journal of Business & Economic Statistics* **4(1)**: pp. 25–38.

Mumtaz, H. & L. Sunder-Plassmann (2010): "Time-varying dynamics of the real exchange rate. A structural VAR analysis." *Bank of England working papers 382*, Bank of England.

Novak, J. (2008): "On the Importance of Accounting Information for Stock Market Efficiency." *Doctoral thesis 134*, Företagsekonomiska institutionen, Uppsala universitet.

Poirier, D. (1995): *Intermediate statistics and econometrics: a comparative approach*. MIT Press.

Primiceri, G. (2005): "Time Varying Structural Vector Autoregressions and Monetary Policy." *Review of Economic Studies* **72(252)**: pp. 821 – 852.

Scheicher, M. (2001): "The comovements of stock markets in Hungary, Poland and the Czech Republic." *International Journal of Finance & Economics* **6(1)**: pp. 27–39.

Strasky, J. (2010): "Can Bayesian econometric methods outperform traditional econometrics in inflation forecasting?" *Master thesis*, Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague.

Sugihara, Y. (2010): "Global contagion of volatilities and volatility risk premiums." *Working paper*.

Tanner, M. A. (1996): *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. Springer Series in Statistics. Springer.

Triantafyllopoulos, K. (2011): "Time-varying vector autoregressive models with stochastic volatility." *Journal of Applied Statistics* **38(2)**.

Tsay, R. S. (2002): *Analysis of financial time series*. John Wiley & Sons, 1st edition.

Wei-Chong, C., S.-C. Loo, L.-B. Ling, & S.-N. Ung (2011): "Return And Volatility Spillover Between Large And Small Stocks In Bursa Malaysia." *International Journal of Business and Social Science* **2(2)**: pp. 176–185.

Appendix A

Outputs from Matlab

This appendix presents additional outputs from Matlab that correspond to particular sections from chapters 5 and 6. Figures A.1 to A.2 present impulse responses of variable orderings (see section 6.1.2), figures A.4 to A.7 show the regression coefficients from the two models described in section 6.2.1.

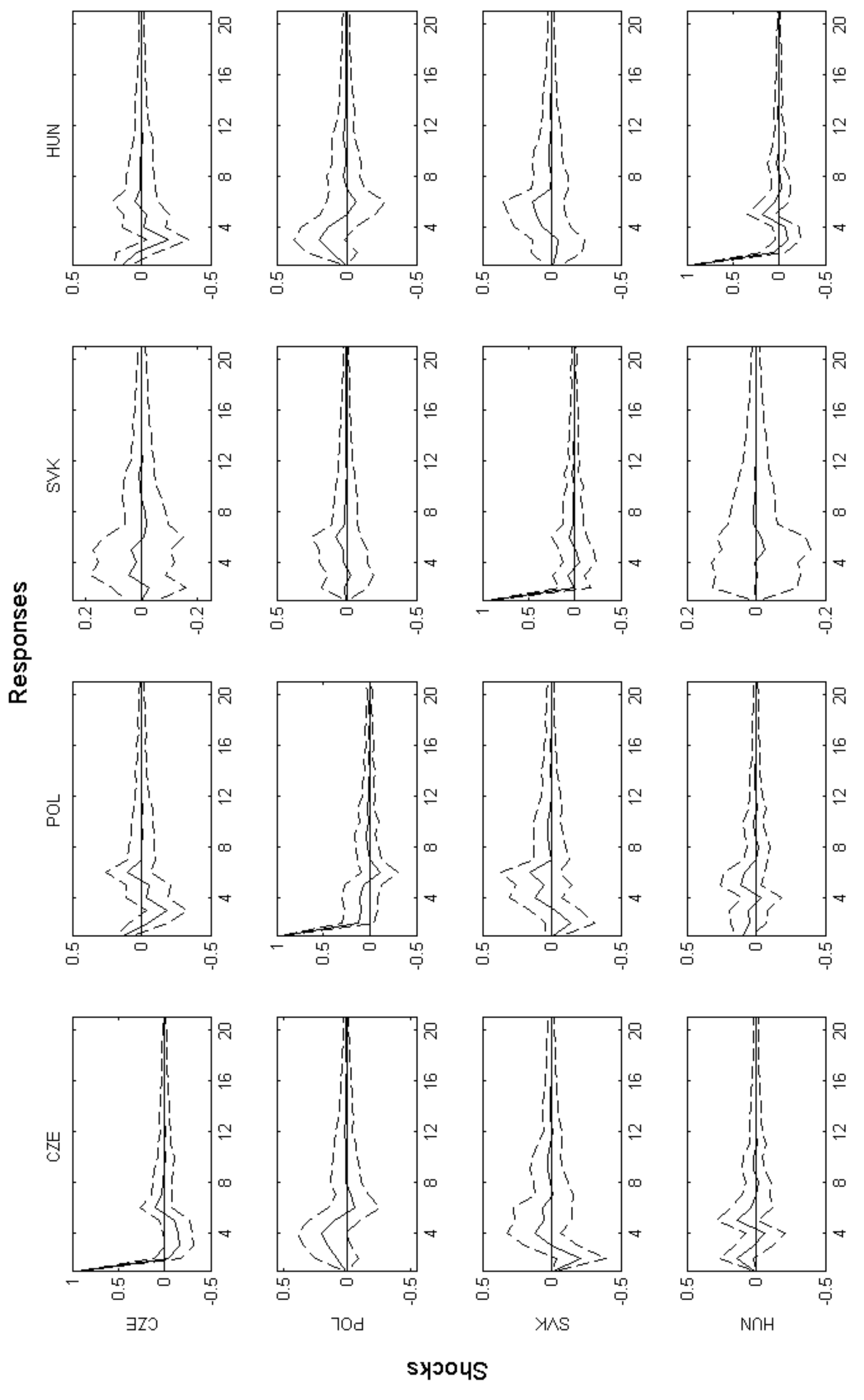

Figure A.1: INW prior, CZE-SVK-HUN-POL

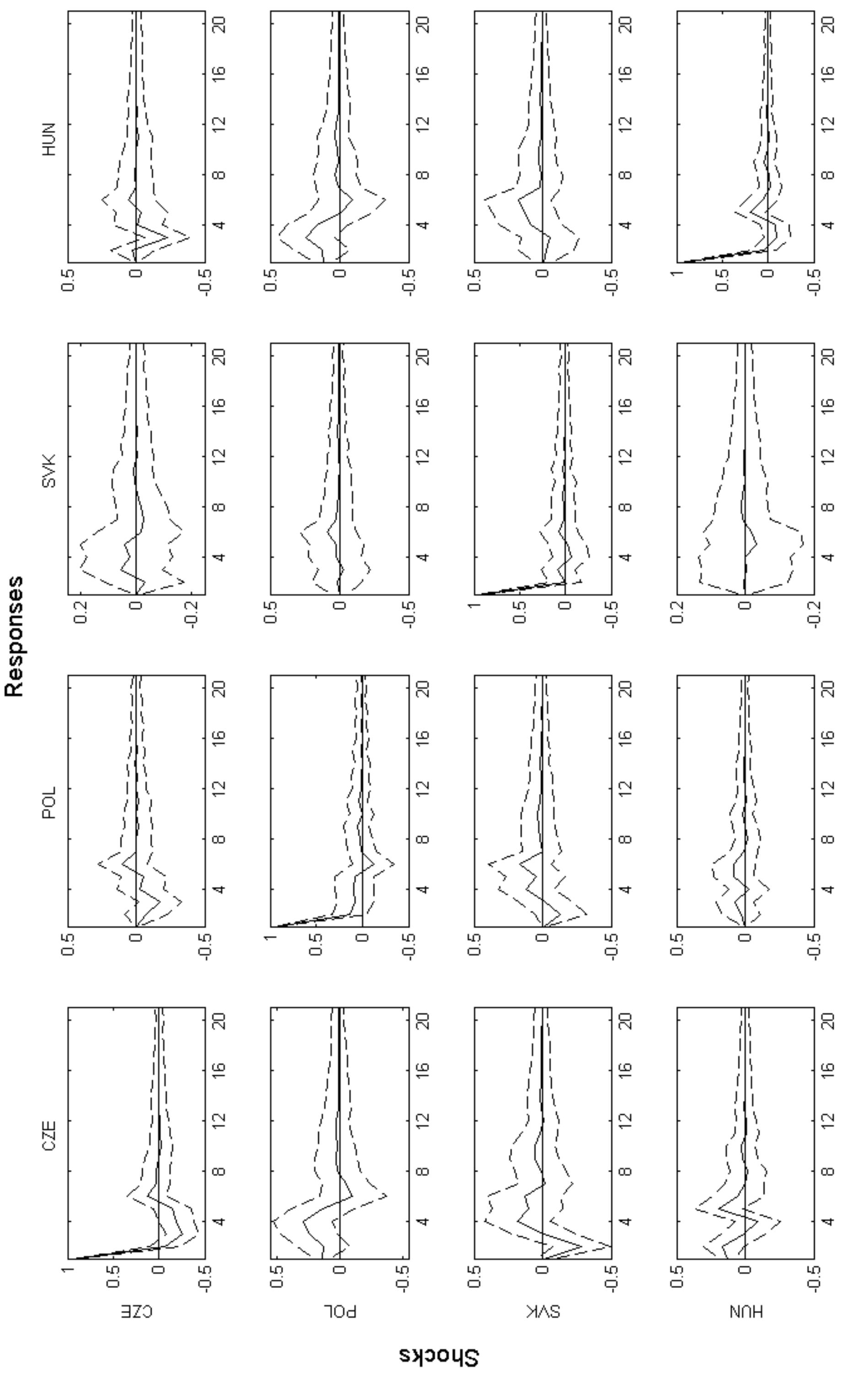

Figure A.2: INW prior, POL-SVK-HUN-CZE

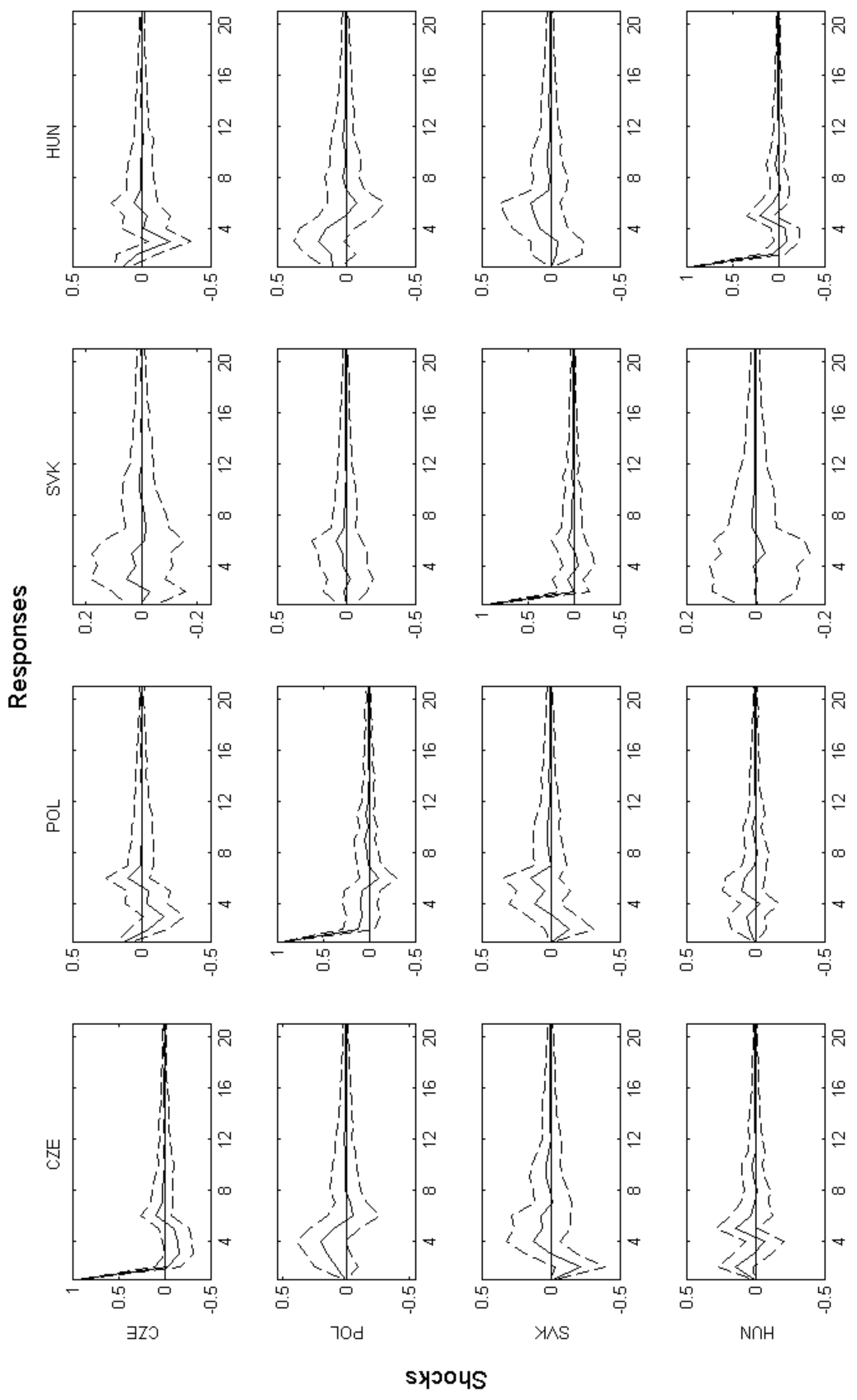

Figure A.3: INW prior, CZE-POL-HUN-SVK

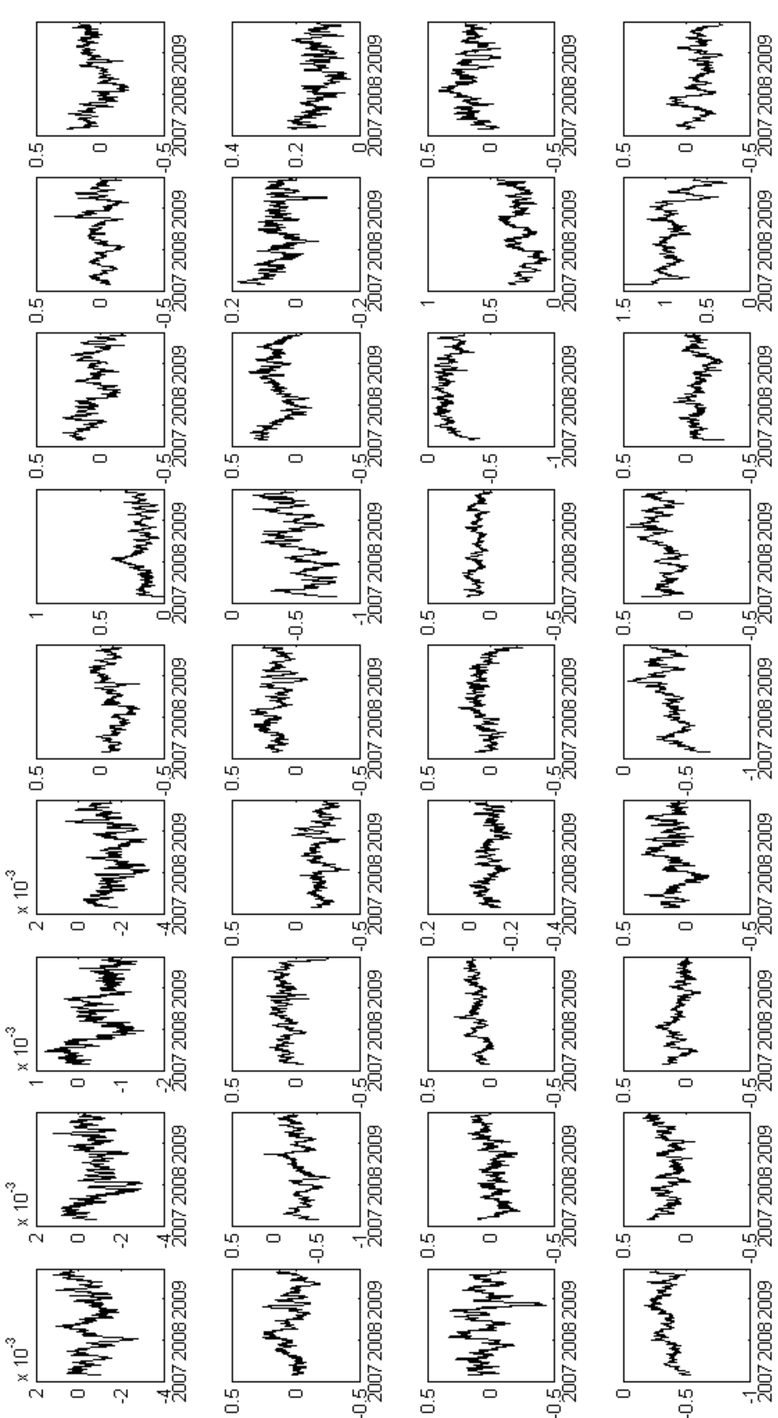

Figure A.4: Robust equation coefficients, lags 1 and 2, model 1

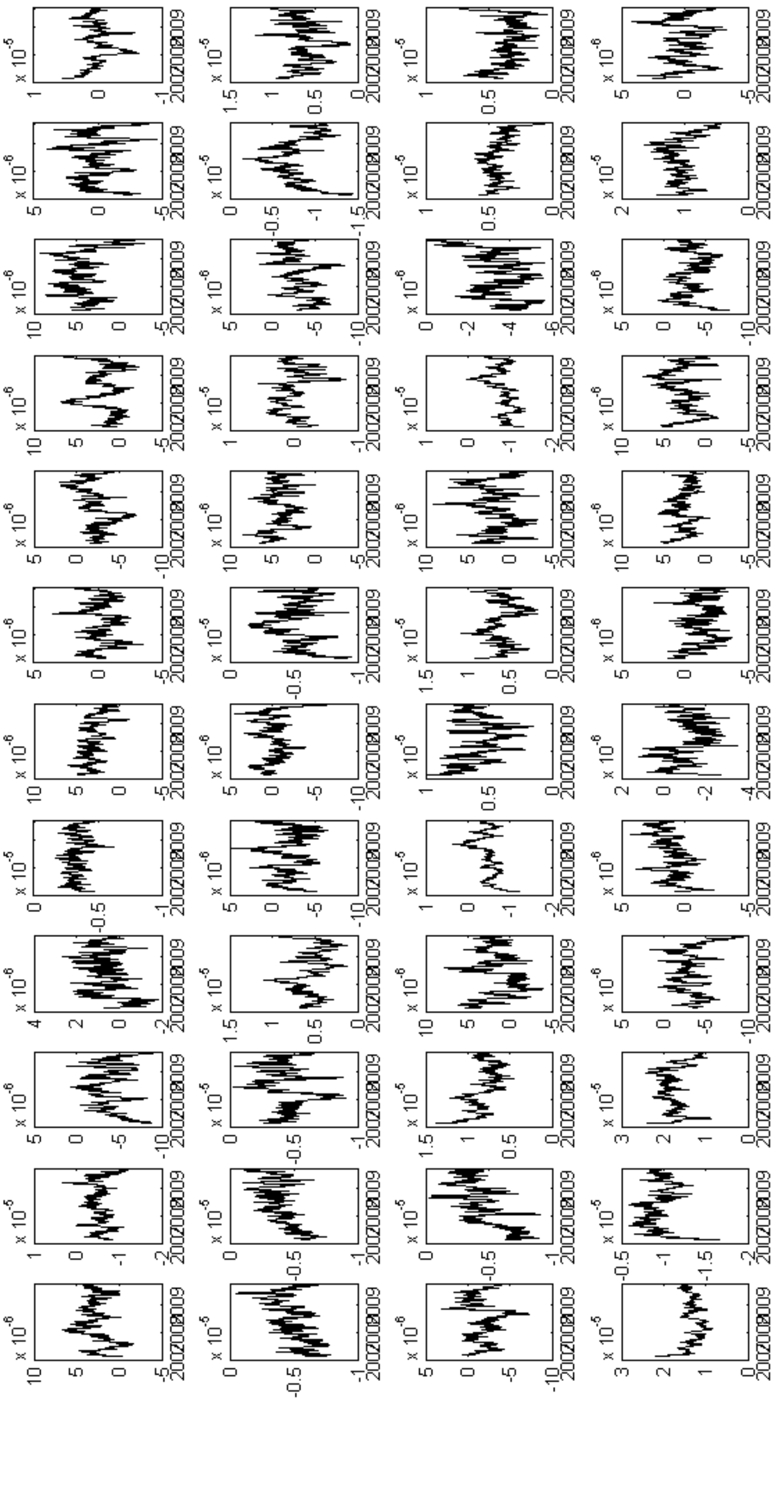

Figure A.5: Robust equation coefficients, lags 3 to 5, model 1

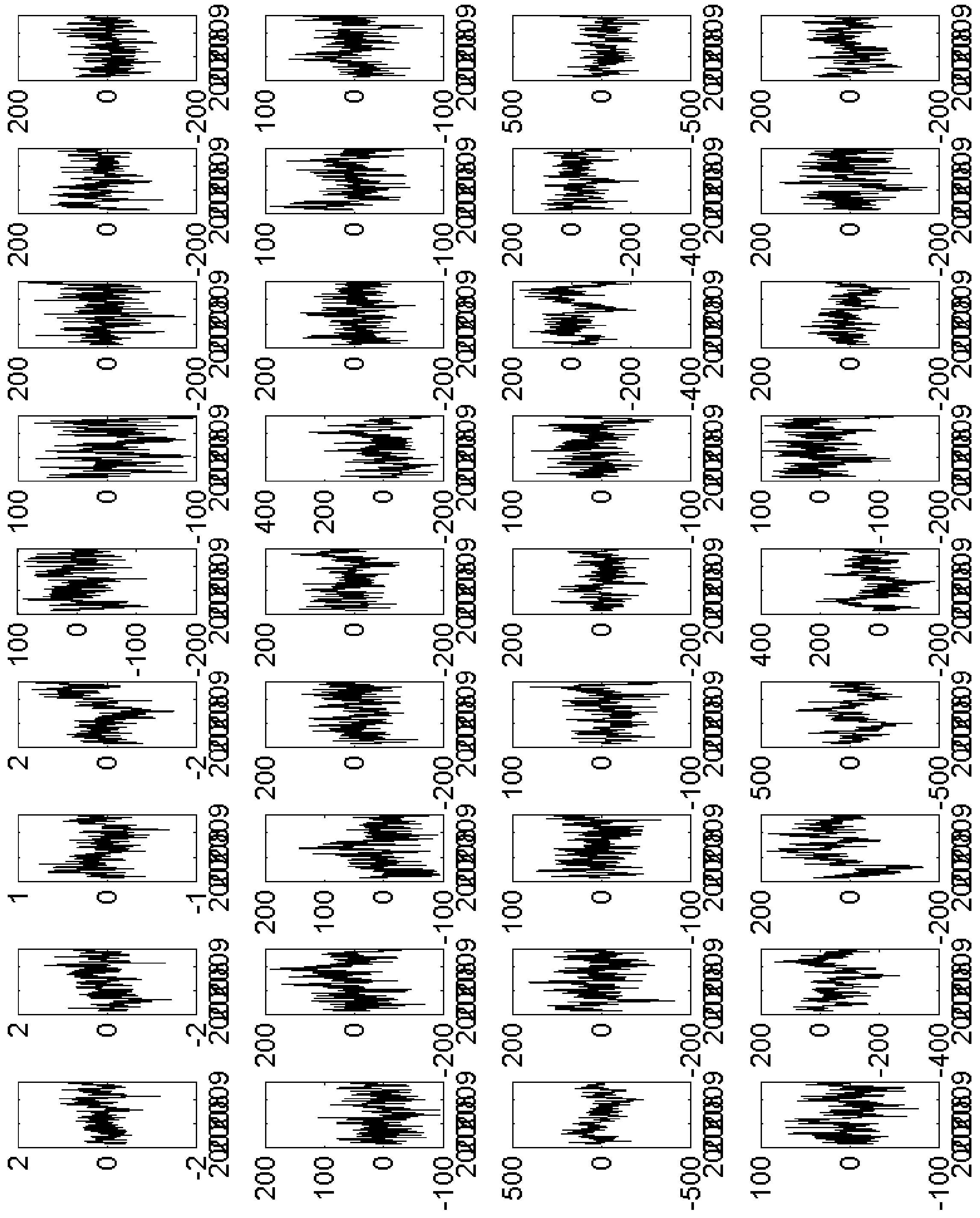

Figure A.6: Robust equation coefficients, lags 1 and 2, model 2

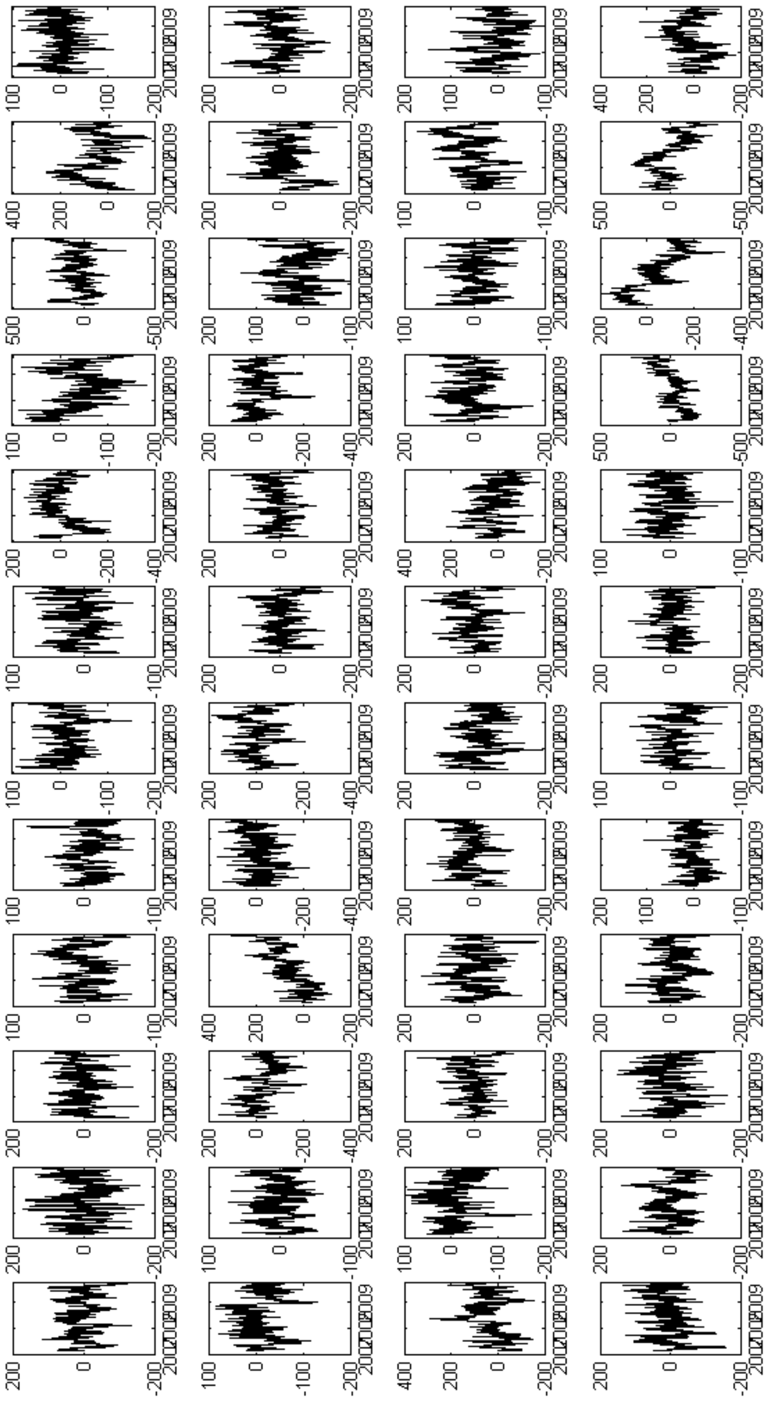

Figure A.7: Robust equation coefficients, lags 3 to 5, model 2