Essais sur la prévision et modélisation d'une économie riche en ressources pétrolières

Oxana Malakhovskaya

université
PARIS-SACLAY

école normale supérieure paris–saclay

NNT : 2019SACLN018

Essays on forecasting and modelling of an energy-based economy

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'École Normale Supérieure Paris-Saclay

Ecole doctorale n°578 Sciences de l'homme et de la société (SHS)
Spécialité de doctorat : Sciences économiques

Thèse présentée et soutenue à Cachan, le 27 mai 2019, par

Mme Oxana Malakhovskaya

Composition du Jury :

M. Franck Portier Professeur, University College London	Président
M. Olivier Darné Professeur, Université de Nantes	Rapporteur
M. Laurent Ferrara Chef du service, Banque de France	Rapporteur
M. Edouard Challe Professeur, École Polytechnique	Examinateur
M. Hubert Kempf Professeur, École Normale Supérieure Paris-Saclay	Directeur de thèse

Thèse de doctorat

Contents

Acknowledgements

I would like to thank many people for their contribution to this dissertation. I express my gratitude to my supervisor, Hubert Kempf for his continuous trust, patience, and support through all the years of research that led to this thesis. I sincerely thank my committee members, Edouard Challe, Olivier Darné, Laurent Ferrara, and Franck Portier, for their insightful comments and suggestions. I am indebted to my co-authors, Alexey Minabutdinov and Boris Demeshev for many hours of joint work and valuable discussions, and their help made a huge difference. I am thankful to my friends and colleagues from the International Laboratory for Macroeconomic Analysis at the National Research University Higher School of Economics for sharing their ideas during research seminars and workshops, and particularly to the director of the laboratory, Sergey Pekarski for invaluable guidance in many aspects of my work, for his support, and for infecting me with incurable love for macroeconomics during my undergraduate studies many years ago.

I thank all the members of the Monetary policy division at the Bank of France where I started working on my thesis as an intern, for sharing their knowledge and research experience and personally Benoit Mojon for creating an excellent working environment and Jean Barthélemy and Magali Marx for answering my numerous questions on DSGE modeling.

I thank all the members of the former Laboratory CES Cachan and personally Emmanuelle Taugourdeau for giving me an opportunity to work in ENS Cachan for several months. I was able to complete a significant part of the first chapter of the thesis there. I thank Warren Ediger for being always available for giving me writing advice.

My special thanks go to my family – my parents, my grandmother, Kristina, André, Alexandre - for their endless love, encouragement, and support over the years and little Angélique and Gabriel for being so cute and making me happier.

List of Figures

List of Tables

General Introduction

After the oil shocks of the 1970s, the role of oil prices in the appearance and development of business fluctuations attracted the attention of the economic community. Hamilton (1983) shows that nine out of ten recessions in the U.S were preceded by oil price shocks and Hamilton (1996) also demonstrates the significant impact of oil prices on the U.S. economy. These studies confirm that the oil price shock can be regarded as a global shock that can trigger a business cycle phase in many countries at the same time. On the other hand, Hooker (1996) challenges this opinion and shows that the severity of oil shocks tends to decrease. In this paper, the author concludes that, unlike the oil shocks of the 1970s, the oil price dynamics of the 1980s did not have a significant impact on the oil sector, and accordingly, the oil price itself has a low effect on key US macroeconomic indicators. For European countries, Mork, Olsen, and Mysen (1994) and Cuñado and Pérez de Gracia (2003) conclude that the real effects of oil price changes have been reduced since the mid-1980s. Using a structural model, Blanchard and Gali (2008) confirm the conclusion of Hooker (1996) and extend it to several developed countries (France, Great Britain, Italy, USA). The authors show that the impact of oil on output and inflation declined over time in these countries. However, the data for Germany and Japan show a different pattern. The reasons for the apparent reduction in the importance of oil shock as a source of business fluctuations are discussed by Baumeister and Peersman (2013). They note that since the 1980s, there has been a reduction in the elasticity of demand in the oil market. Thus, the same price increase caused by the negative supply shock causes a smaller drop in world production and therefore has less impact on macroindicators.

As this short review shows, the impact of the oil shock on the U.S. economy and other developed countries, mainly oil importers, is frequently studied. However, the impact of oil market shocks on emerging and oil-exporting countries is rarely analysed in academic literature.

An example of a study of the world natural resources price shocks effect on an exporting economy is the paper by Charnavoki and Dolado (2014). Using Canadian data, the authors show that to a large extent the resources price fluctuations are determined by global demand shocks and specific resource supply shocks. The authors conclude that a positive resources price shock has a positive effect on the current account regardless of the type of shock.

At the same time, the presumptive influence of oil prices on an oil-exporting economy may be significant and highly different than on an oil-importing economy. For example, in economic literature and mass media it is generally agreed that economic activity in Russia crucially depends on oil price dynamics. This perception is based on the fact that Russia is one of the world's largest oil producers, with oil and gas exports amounting to 60% of all export revenues. However, quantitative estimates of oil price effects are extremely scarce. The first chapter of this thesis fills this gap, estimating an oil price effect on the Russian economy in a general equilibrium framework.

The goal of the chapter is twofold. The first is to elaborate a general-equilibrium model - suitable for structural analysis and forecasting - with some specific features of commodity-exporting countries. The second goal is to identify the main sources of volatility of key macroeconomic variables in Russia and answer the question that we raised in the title of the chapter: are commodity prices important as a factor of business cycles in an export-oriented economy?

Like in many other DSGE models, such as Smets and Wouters (2003), the DSGE model presented in the thesis is estimated by Bayesian techniques. Although DSGE models can generally be estimated by frequentist methods, the Bayesian approach has recently gained popularity.

Bayesian estimation is a combination of maximum likelihood estimates (determined by the structure of the model and the data) with some prior knowledge described with prior distributions in order to construct a posterior distribution for the parameters of interest. Certainly, the use of prior information may raise questions about the origin of the prior knowledge and its credibility. However, from a practical point of view, using prior distributions improves estimates of parameters. Pre-sample information is particularly necessary when one deals with emerging economies . When the sample size is limited, the maximum likelihood function is often almost flat, and its combination with some reasonable non-flat prior can help achieve identification (Fernandez-Villaverde (2010)). A similar problem may emerge when a medium- or a large-size model is to be estimated even for a developed economy as the necessary amount of data for such a model is not available. It explains the popularity of the Bayesian approach in estimation and estimation of DSGE model in large extent. Besides, the advances in Bayesian theory make an expanding set of tools available for researchers to estimate and evaluate their models (Guerron and Nason (2013)).

Our model yields plausible estimates, and the impulse response functions are in line with empirical evidence. However, even a better test of a good model performance would be its consistent forecasting accuracy. Besides, the particularities of an energy-based economy may change a ranking of good forecasting models with respect to a developed and non energy-based economy. It motivates the three next chapters devoted to forecasting with structural as well as non-structural models.

Accurate forecasts are crucial for timely and correct macroeconomic policy due to policy lags and statistical data-availability delays. Currently, a widely used tool for doing macroeconomic forecasting is a vector autoregression (VAR) model and its modifications. VAR models have become a workhorse in forecasting applications following a critique aimed at Keynesian-type models with systems of equations (SOE) that were widely spread previously for structural and non-structural analysis. Contrary to SOE models, a VAR model is a dynamic plainly-specified model without any additional restrictions on the joint dynamics of endogenous variables. Nevertheless, a correct representation of the actual time series dynamics requires a sufficient amount of lags. Moreover, the correct representation of the information set of a central bank requires many endogenous variables. However, an increasing number of lags and endogenous variables in an unrestricted VAR model bear the risk of the over-parametrization, non-efficient estimation, and high forecast errors. To circumvent this problem it is possible to shrink the parameters by imposing some prior distributions and to estimate a Bayesian VAR (BVAR), while for a small-size VAR, it is usually sufficient to apply the OLS to each equation consecutively (frequentist VAR).

Although the use of BVARs has become very popular in research papers recently, the reviews of this approach are scarce with some prominent exceptions (Karlsson (2012), Del Negro and Schorfheide (2011), Canova (2007)) that are mathematically intensive and hardly comprehensible for novices in Bayesian analysis. Moreover, most of the reviews are not complemented by a guide to estimate a BVAR in econometric software. Finally, no review provides the reader with a detailed algorithm of forecasting with BVAR though the estimation of the BVAR in a reduced form is usually done for forecasting purposes.

The use of Bayesian techniques and their application for estimation and forecasting with a reduced-form BVAR model is presented in the **second chapter** of the thesis. The first part of the chapter is devoted to the classification of the prior distributions that are frequently used in macroeconomic applications. We show how the parameters of posterior distributions can be computed for these priors. In the second part of the chapter the point and density forecasting with BVAR is described in detail. The chapter does not contain any empirical application and is considered as purely methodological. To make the chapter more useful for readers, it is complemented by a package in R with the same notations as in this chapter.

The methods reviewed in the second chapter are applied in the **third chapter** where we evaluate the forecast performance of Bayesian vector autoregressions (BVARs) on Russian data. As indicated earlier, central banks monitor many time series to formulate their policies. It leads to a generally agreed idea that a potentially good forecasting model should be able to take into account a large sample of series not to lose relevant information. To address the issue for an energy-based economy, we estimate BVARs of different sizes and compare the accuracy of their out-of-sample forecasts with those obtained with unrestricted vector autoregressions, random walk with drift, and a set of univariate models. We have two goals. First, we compare the forecast accuracy of BVAR models with the competing ones for 23 important macroeconomic indicators. Second, we question if a high-dimensional model is always better than a low-dimensional one in terms of forecast accuracy.

The second goal is motivated by a strand of research that generally confirms the hypothesis that increasing the number of variables included in the BVAR model helps to forecast more accurately (Beauchemin and Zaman (2011), Bloor and Matheson (2010), Bańbura, Giannone, and Reichlin (2010) etc). However, all indicated studies used data of a developed economy, so their conclusions cannot be taken for granted if applied to an emerging energy-based economy.

Our analysis delivers two meaningful results. First, we show that many Russian macroeconomic indicators can be forecast by BVARs more accurately than by competing models. Second, contrary to several other studies, we do not confirm that the relative forecast error monotonically decreases with increasing the cross-sectional dimension of the sample. In half of those cases where a BVAR appears to be the most accurate model, a small-dimensional BVAR outperforms its high-dimensional counterpart.

While the third chapter of the thesis considers non-structural models only, the **forth chapter** compares non-structural forecasts - obtained with BVAR and VAR models - to those made with a structural DSGE model presented in the first chapter of the dissertation. As indicated above, modelling an economy in a DSGE framework has become a mainstream tendency for researchers from academic community and central banks over the past 15 years due to substantial advances in estimation methods of this kind of models. An essential advantage of the DSGE models is that they are microfounded and therefore not prone to Lucas critique. However, the DSGE models remain a stylized description of the reality. The question we address in the chapter is whether the structural DSGE models can be so efficient in terms of forecast accuracy as nonstructural vector autoregressions are. There is no general consensus about a relative forecast performance of DSGE models with respect to non-structural VARs. For example, Smets and Wouters (2007) show that a new Keynesian DSGE model with a sufficient amount of rigidities can forecast not worse (and even better in some cases) than BVARs. In contrast, Edge and Gürkaynak (2011) claim that forecasting accuracy of different models, including both BVAR and DSGE, is low. Both papers were estimated on the US data. As far as oil-exporting economies are concerned, changes in the price of oil in the global market may be considered as a potential source of business cycles in these countries and consequently may contain information relevant for forecasting purposes of their macroeconomic indicators. A priori, it is not known which model - structural or non-structural - gains at a higher level from taking into account the oil price explicitly, that motivates the study presented in the forth chapter of the dissertation. The empirical application is done using Russian data. On the basis of mean-square forecast errors (MSFE), we conclude that the DSGE model is usually inferior to BVAR model in terms of forecasting accuracy but the difference is not too large. At the same time, DSGE model allows the user to obtain a forecast with the minimal MSFE for some variables and some forecast horizons considered.

The fifth chapter of the thesis extends the forth one in terms of the tool applied and extends the first one in terms of the research question. This chapter constructs and estimates a structural Bayesian VAR (SBVAR) model to study the effects of different oil market shocks on Russian economy. The first chapter of the dissertation assumes that the dynamics of the oil revenues is completely exogenous. It is justified while the analysis is done in the general equilibrium framework but it seems reasonable to reject this assumption in the context of the partial equilibrium analysis.

In the literature, the irrelevance of the exogeneity assumption of the price of oil for large importing countries was suggested by Barsky and Kilian (2002) and Barsky and Kilian (2004). They explain the need for a separate assessment of the supply of oil and demand of oil shocks effects. If these shocks are analysed separately, the empirical studies demonstrate that the role of supply shocks has declined over time (Edelstein and Kilian (2007), Herrera and Pesavento (2009)). A seminal paper by Kilian (2009) identifies three different types of shocks in the oil market and shows that the variations of the price of oil affect the U.S. economy in different ways in dependence of the type of the shock that resulted in the price variation. The Kilian's paper sparked interest of the academics to the oil markets shocks effects (Lippi and Nobili (2012), Baumeister and Peersman (2013), Aastveit (2014), Baumeister and Hamilton (2015c) Stock and Watson (2016), and others) but all of them with very few exceptions (Cavalcanti and Jalles (2013), Charnavoki and Dolado (2014)) focused their attention of the oil market shocks effects on developed oil importing countries. Cavalcanti and Jalles (2013) study the effects of oil market shocks on the Brazilian economy (emerging oil importing country), Charnavoki and Dolado (2014) study the effects of commodity market shocks on the Canadian economy (developed copper exporting economy). However the the academic literature is still almost unaware about the effects of the oil price shocks on an oil export oriented economy. The fifth chapter fills the gap and quantifies the effects of different kinds of oil market shocks on several macroeconomic indicators. From the methodological point of view, the chapter continues a strand of research about appropriate identification schemes in SVAR models. We use a new method of imposing sign restrictions proposed by Baumeister and Hamilton (2015b).

Pioneer papers on oil market shocks effects exploited the recursive identification scheme (for example, Kilian (2009) and Blanchard and Gali (2008)) but due to widely known disadvantages of the triangular scheme, the sign restrictions gained the ground shortly after in papers by Peersman and Van Robays (2009), Baumeister, Peersman, and Van Robays (2010), Kilian and Murphy (2012), and others. In its turn, a crucial critique of the latter identification scheme presentted by Baumeister and Hamilton (2015b) is based on a idea that in case of traditional sign restrictions some functions of estimated parameters of a SVAR model within their identifying set - including the impulse response functions - may be governed completely by an implicit prior imposed by a researcher and not by the dataset. Baumeister and Hamilton (2015b) suggest that a researcher uses an explicit prior distributions for coefficients that may have some economic meaning (e.g., elasticity of supply and demand) and they show some empirical applications of this scheme in to identify labour market shocks (Baumeister and Hamilton (2015b)), monetary policy shocks (Baumeister and Hamilton (2015a)) and oil market shocks (Baumeister and Hamilton (2015c)). The model pesented in the fifth chapter of the thesis extends the analysis by Baumeister and Hamilton (2015c)) to study the effects of shocks identified with explicit sign restrictions on key Russian macroeconomic indicators. To the best of the knowledge of the author, no paper so far has shown how the shocks identified with explicit sign restrictions may affect variables external to the oil market. Besides, the chapter is a rare example of econometric analysis of oil market shocks effects on the Russian economy, and the first one using the SBVAR model. The estimation shows mixed results about the effects of oil market shocks on the real monetary incomes and CPI inflation but without any reservations indicates that two of three oil demand shocks considered in the study positively influence the industrial production index. The quantitative estimates of the effects on the industrial production are surprisingly high and may be challenged again in the future research.

Chapter 1

Are Commodity Price Shocks Important? A Bayesian Estimation of a DSGE Model for Russia[1]

1.1 Introduction

In the economic literature, there is a widespread belief that economic activity in Russia crucially depends on oil price dynamics. This perception is based on the fact that Russia is one of the world's largest oil producers, with oil and gas exports amounting to \$342 bln in 2011, accounting for 18.5% of Russian GDP and one-half of federal budget revenues. In this situation, it seems evident that oil price shocks could dominate Russian business cycles and long-run dynamics of macroeconomic variables. However, quantitative estimates of oil price effects are scarce. For example, Rautava (2002) analyzes the impact of oil prices on the Russian economy using the VAR methodology and cointegration techniques and discovers that, in the long run, a 10% increase in oil prices is associated with a 2.2% growth in Russian GDP. Their sample covered the period from Q1 1995 to Q3 2001. Jin (2008) uses a similar methodology and claims that in the 2000s, a permanent 10% increase in oil prices was associated with a 5.16% growth in Russian GDP. In both papers, the authors use quarterly data, so the time series seem to be too short for cointegration analysis to

[1] co-authored with Alexey Minabutdinov, NRU HSE; published as Malakhovskaya and Minabutdinov (2014)

have good estimation properties. Moreover, neither of these papers raises questions about the short-run impact of oil prices on macroeconomic variables and the role of oil prices as a potential factor of the business cycle.

Since the 1990s, there has been a growing interest in Dynamic Stochastic General Equilibrium (DSGE) models for macroeconomic analysis from both academia and central banks. Contrary to VAR, DSGE models provide a theoretical explanation of different interdependencies among variables in the economy. These models allow to determine the factors of business cycles, forecast macroeconomic variables, identify the impact of structural changes, etc. Sosunov and Zamulin (2007) analyze an optimal monetary policy in an economy sick with Dutch disease in a general equilibrium framework. They calibrate their model on Russian data, but they assume that the shock to the terms of trade is the only source of uncertainty in the economy, and they do not consider the relative importance of this kind of shock in real data. Semko (2013) estimates a modified version of the model by Dib (2008) using Russian data with a focus on optimal monetary policy. He mentions that his results indicate that the impact of oil price shock on GDP is small, as a rise in output in the oil production sector is associated with an output decline in manufacturing and non-tradable sectors, but quantitative estimates of the impact are not provided in the paper.

The purpose of our paper is twofold. The first goal is to elaborate a theoretical model with a special focus on commodity-exporting countries that is suitable as a basis for policy implications. The second goal is to determine the main sources of volatility of key macroeconomic variables in Russia and answer the question that we raised in the title of the paper: are commodity prices important as a factor of business cycles in an export-oriented economy?

Our paper has some policy implications. The belief that economic activity in Russia is mostly determined by oil price dynamics was an argument for the exchange rate management policy. Recently the Central Bank of Russia announced a new course of monetary policy based on an inflation targeting policy from 2015 onwards. It is crucial to understand what role commodity exports play in business cycles in order to assess the potential success of this policy switch. While the traditional Mundell-Fleming model states that flexible exchange rates dominate fixed exchange rates if foreign real shocks prevail, this prescription is called into doubt when an adjustment requires a substantial devaluation or revaluation of exchange rates (Cespedes, Chang, and Velasco (2004)). In this case, an exchange rate peg may be desirable.

In this paper, we modify the Kollmann's model (Kollmann (2001)) and assume external habit formation, a cashless economy, and CRRA (constant rate of relative aversion) preferences of households as in Smets and Wouters (2003) and Dam and Linaa (2005). The model contains a number of real and nominal frictions, like sticky prices and wages, local currency pricing, and capital adjustment costs. It is known from previous research that rigidities play a key role in the fitting and forecasting performance of DSGE models (Christiano, Eichembaum, and Evans (2005), Smets and Wouters (2007)). Additionally, we assume that the nominal interest rate is an instrument of monetary policy and increase the number of structural shocks under consideration. We introduce ten structural shocks. Nine of them are relatively standard, while the tenth is a commodity export shock. Next, we estimate the model on Russian data using Bayesian methods. Our results show that, while this shock contributes a lot to GDP variation, the most important factors of business cycles in Russia are domestically based.

We proceed as follows: Section 2 presents the model. For the sake of convenience, we present the full set of equations. In Section 3, we review our estimation techniques and discuss our results. Section 4 concludes.

1.2 Model

In this section, we present the model that we estimate in the next section. We assume two types of firms that produce intermediate and final goods. The final sector is competitive, and intermediate sector is monopolistic competitive. Households can own capital and rent it, as well as labor services to intermediate goods firms. They can optimize both intertemporally and intratemporally. Prices and wages are rigid due to a mechanism à la Calvo. A final good can be used for consumption and for investment. The final good is aggregated from domestic and imported intermediate ones. Export and import are possible only for intermediate products and are priced in local currency. Financial markets are incomplete and households can own domestic and foreign bonds (or issue debt).

The core of our model is that by Kollmann (2001)[2] but we have made some important modifications. First of all, we assume external habit formation, a cashless economy and CRRA preferences. Secondly, we include revenues from oil exports which are assumed to increase households' wealth exogenously. Finally, we assume that monetary policy follows an interest rate rule.

1.2.1 Production sector

Final goods production

We assume that the only final good is produced by combining intermediate domestic and imported aggregates using Cobb–Douglas technology:

$$Q_t = \left(\frac{1}{\alpha_d} Q_t^d\right)^{\alpha_d} \left(\frac{1}{\alpha_{im}} Q_t^{im}\right)^{\alpha_{im}}, \quad 0 < \alpha_d < 1, \alpha_{im} = 1 - \alpha_d \tag{1.1}$$

Q_t denotes the final output index. Q_t^d and Q_t^{im} are indices of aggregate domestic and foreign intermediate goods production, respectively, and they are defined as Dixit–Stiglitz aggregates:

$$Q_t^d = \left(\int_0^1 q_t^d(j)^{\frac{1}{1+\upsilon_t}} dj\right)^{1+\upsilon_t} \quad Q_t^{im} = \left(\int_0^1 q_t^{im}(j)^{\frac{1}{1+\upsilon_t}} dj\right)^{1+\upsilon_t} \tag{1.2}$$

where $q_t^d(j)$ and $q_t^{im}(j)$ are quantities of type j intermediate goods produced domestically and abroad, respectively, and sold on domestic market, and υ_t is a random net mark-up rate. In other words, in the intermediate goods market, there is a continuum (of unit measure) of producers, and we use index j to indicate them. Each producer sells her own variety (also indicated by j) in the monopolistic competitive market. The final sector is perfectly competitive and does not incur any cost above the value of the intermediate bundles.

A cost-minimization problem for the final producer can be written as:

$$\min TC_{final} = \int_0^1 p_t^d(j) q_t^d(j) dj + \int_0^1 p_t^{im}(j) q_t^{im}(j) dj \tag{1.3}$$

subject to constraints (1.1) and(1.2) where $p_t^d(j)$ and $p_t^d(j)$ represent prices of domestic and imported type j intermediate products respectively, both expressed in domestic currency.

[2]Our notations are close to those of Dam and Linaa (2005), whose model is also a modification of Kollmann (2001).

The demand functions for any variety (domestic or imported) of intermediate products as well as for intermediate aggregates are derived as a solution of the cost-minimization problem. They are given by:

$$q_t^d = Q_t^d \left(\frac{p_t^d(j)}{P_t^d}\right)^{-\frac{1+\upsilon_t}{\upsilon_t}} \qquad q_t^{im} = Q_t^{im} \left(\frac{p_t^{im}(j)}{P_t^{im}}\right)^{-\frac{1+\upsilon_t}{\upsilon_t}} \tag{1.4}$$

and

$$Q_t^d = \alpha_d \frac{P_t}{P_t^d} Q_t \qquad Q_t^{im} = \alpha_{im} \frac{P_t}{P_t^{im}} Q_t \tag{1.5}$$

letting P_t^d and P_t^{im} be the price indices of intermediate domestic and foreign bundles sold in the domestic market, respectively, and P_t representing the final good price index. We postulate that intermediate goods are packed in a bundle at no cost, and the value of a bundle is equal to the value of its ingredients. The total revenue of the final producers is equal to their total costs as they are competitors and operate on a zero-profit bound. This means that:

$$P_t^d Q_t^d = \int_0^1 p_t^d(j) q_t^d(j) dj \qquad P_t^{im} Q_t^{im} = \int_0^1 p_t^{im}(j) q_t^{im}(j) dj \tag{1.6}$$

So we get:

$$P_t^d = \left(\int_0^1 p_t^d(j)^{-\frac{1}{\upsilon_t}} dj\right)^{-\upsilon_t} \qquad P_t^{im} = \left(\int_0^1 p_t^{im}(j)^{-\frac{1}{\upsilon_t}} dj\right)^{-\upsilon_t} \tag{1.7}$$

A zero-profit condition for the final good sector requires:

$$P_t^d Q_t^d + P_t^{im} Q_t^{im} = P_t Q_t \tag{1.8}$$

Hence, the final good price index is determined by a weighted geometric mean of domestic and imported aggregates price indices:

$$P_t = \left(P_t^d\right)^{\alpha_d} \left(P_t^{im}\right)^{\alpha_{im}} \tag{1.9}$$

Intermediate sector

An intermediate good j is produced from labor and capital with Cobb–Douglas technology:

$$y_t(j) = A_t K_t(j)^{\psi} L_t(j)^{1-\psi}, \qquad \text{where } 0 < \psi < 1 \tag{1.10}$$

where $y_t(j)$ is an output of an intermediate type j firm, A_t is a technology parameter, $K_t(j)$ is capital stock that firm j holds (capital utilization is assumed to be equal to one), and $L_t(j)$ is the amount of labor services utilized by firm j and represents a Dixit–Stiglitz aggregate of different varieties of labor services provided by households:

$$L_t(j) = \left(\int_0^1 l_t\,(h, j)^{\frac{1}{1+\gamma}}\, dh \right)^{1+\gamma} \tag{1.11}$$

where $l_t(h, j)$ is the amount of labor services of household h employed by firm j. Here we assume that there is a continuum (of unit mass) of households (indexed by h), their labor services are differentiated, and the labor market is monopolistic competitive. So each household is a monopolistic supplier of its labor and sets the wage on its own (we describe the mechanism of wage-setting below). On the contrary, capital is homogenous. The law of motion of the technology process is declared below. This, the total costs of firm j are the following:

$$TC_t(j) = R_t^K K_t(j) + \int_0^1 w_t(h) l_t(h, j) dh, \tag{1.12}$$

where R_t^K is the rental rate of capital, and $w_t(h)$ is the wage of household h. The problem of an intermediate firm consists in minimizing $TC_t(j)$ s.t. (1.10). The first-order conditions imply that demand functions for aggregate labor and capital can be written as:

$$L_t(j) = \frac{y_t(j)}{A_t} \left(\frac{\psi}{1-\psi} \cdot \frac{W_t}{R_t^K} \right)^{-\psi} \tag{1.13}$$

$$K_t(j) = \frac{y_t(j)}{A_t} \left(\frac{\psi}{1-\psi} \cdot \frac{W_t}{R_t^K} \right)^{1-\psi} \tag{1.14}$$

Additionally,

$$l_t(h, j) = L_t(j) \left(\frac{w(h)}{W_t} \right)^{-\frac{1+\gamma}{\gamma}} \tag{1.15}$$

As far as the total labor costs for intermediate firm j are concerned, they are equal to labor expenses for all varieties:

$$W_t L_t(j) = \int_0^1 w_t(h) l_t(h, j) dh, \tag{1.16}$$

the aggregate wage index is

$$W_t = \left(\int_0^1 w_t(h)^{-\frac{1}{\gamma}} dh \right)^{-\gamma} \tag{1.17}$$

The marginal cost of firm j is equal to:

$$MC_t(j) = A_t^{-1} W_t^{1-\psi} R_t^{K\psi} \psi^{-\psi} (1-\psi)^{\psi-1} \tag{1.18}$$

Therefore, the marginal cost is the same for all firms in the market; it allows us to omit an index of a firm in what follows. Moreover, the total cost is a linear function of output, and the marginal cost is independent of output. This lets us consider problems of setting domestic and export prices separately. We assume that intermediate goods are sold on domestic and international markets:

$$y_t(j) = q_t^d(j) + q_t^{ex}(j), \tag{1.19}$$

where $q_t^d(j)$ and $q_t^{ex}(j)$ are quantities of intermediate good j sold on the domestic market and exported, respectively. All the intermediate goods sold in the domestic market are bought by the final producer. We postulate that intermediate firms can practice price discrimination between domestic and foreign markets. In general, this means that:

$$S_t p_t^{ex}(j) \neq p_t^d(j) \tag{1.20}$$

where $p_t^d(j)$ and $p_t^{ex}(j)$ are price indices of intermediate good j sold in the domestic market and exported, respectively, and S_t is a nominal exchange rate (expressed as a domestic currency price of foreign currency). The assumption about price discrimination and, consequently, the violation of the law of one price is motivated by a great number of theoretical and empirical papers (see, for example, Balassa (1964), Taylor and Taylor (2004)) which show that the absolute purchasing power parity (PPP) does not hold, at least, in the short-run. In the new open economy macroeconomics literature there are several microfounded approaches to model deviations from the PPP, and Ahmad, Lo, and Mykhaylova (2011) offer a very good review of them.[3] In this paper, we assume that intermediate firms – both domestic and foreign – and households carry out staggered price and wage setting, respectively, and the exporting and importing activity is characterized by price-to-market behavior (Knetter (1993)). This means that the prices are set in the

[3]According to Ahmad et al. (2011), there are four approaches: presence of both tradables and non-tradables (e.g., Corsetti, Dedola, and Leduc (2008)), home bias in consumption (e. g., Faia and Monacelli (2008)), price rigidity (e.g., Bergin and Feenstra (2001)), and local currency pricing (Chari, Kehoe, and E.McGrattan (2002)).

local (buyer's) currency. The staggered price and wage setting is implemented à la Calvo (Calvo (1983)). The probability of a price-changing signal is equal to $1-\theta_d$. Because the number of firms is large, in accordance with the law of large numbers, we can define the share of firms reoptimizing their prices each period as equal to $1-\theta_d$, as well. All the firms are obliged to meet the demand for their products at the set price. Suppose a firm gets a signal and is allowed to adjust its price. In this case, the price chosen by the producer is one that maximizes an expected discounted flow of her future profits:

$$\widetilde{p}_t^d(j) = \underset{p_t^d(j)}{\arg\max}\, E_t \left[\sum_{\tau=0}^{\infty} \theta_d^\tau \lambda_{t,t+\tau} \Pi_{t+\tau}^{d,j} \left(p_t^d(j) \right) \right] \tag{1.21}$$

where $\widetilde{p}_t^d$ is a reset price; $\Pi_{t+\tau}^{d,j}$ is the profit of intermediate firm j from selling its product in the domestic market (superscript d) at time $t+\tau$; $\lambda_{t,t+\tau}$ is a stochastic discount factor of nominal income (pricing kernel). It is assumed to be equal to the intertemporal marginal rate of substitution in consumption between periods t and $t+\tau$ and is given by:

$$\lambda_{t,t+\tau} \equiv \beta^\tau \frac{U'_{C,t+\tau}}{U'_{C,t}} \cdot \frac{P_t}{P_{t+\tau}} \tag{1.22}$$

While solving its problem of profit maximization, the firm takes into account all the expected profits until the next price-changing signal comes. As the number of periods to be taken into account is not known in advance, the producer maximizes her discounted profit over an infinite horizon, and each profit is multiplied by the probability that the firm has not received a new price-changing signal before. The instantaneous profit of intermediate producer j from selling her variety in the domestic market is defined as:

$$\Pi_t^{d,j} = \left(p_t^d(j) - MC_t \right) q_t^d(j) = \left(p_t^d(j) - MC_t \right) \left(\frac{p_t^d(j)}{P_t^d} \right)^{-\frac{1+\upsilon_t}{\upsilon_t}} Q_t^d \tag{1.23}$$

Therefore, the problem facing the producer is to maximize (1.21) subject to (1.23). The first order conditions result in the following equation for the optimal price:

$$E_t \sum_{\tau=0}^{\infty} \theta_d^\tau \lambda_{t,t+\tau} \frac{1}{\upsilon_{t+\tau}} (P_{t+\tau}^d)^{\frac{1+\upsilon_{t+\tau}}{\upsilon_{t+\tau}}} Q_{t+\tau}^d \widetilde{p}_t^d(j)^{-\frac{1+\upsilon_t}{\upsilon_t}-1} \times \\ \times \left(\widetilde{p}_t^d(j) - (1+\upsilon_{t+\tau}) MC_{t+\tau} \right) = 0 \tag{1.24}$$

1.2.2 Foreign Sector

Export

We assume that the structure of a foreign economy is the same as the structure of a domestic one. Similar to the demand for domestic intermediate goods, the export demand is assumed to be defined as:

$$Q_t^{ex} = \alpha_{ex} \left(\frac{P_t^{ex}}{P_t^f} \right)^{-\eta} Y_t^f \tag{1.25}$$

where P_t^{ex} is the price index of the intermediate domestic bundle exported abroad, P_t^f is an aggregate price level in the foreign economy, and Y_t^f is a quantity of final goods produced in the foreign economy. Both prices are expressed in foreign currency. Similar to the demand for a particular type of intermediate goods in the domestic economy, export demand for a variety j ($q_t^{ex}(j)$) is given by:

$$q_t^{ex}(j) = Q_t^{ex} \left(\frac{p_t^{ex}(j)}{P_t^{ex}} \right)^{-\frac{1+\upsilon_t}{\upsilon_t}} \tag{1.26}$$

with the same elasticity of substitution that characterizes the domestic demand:

$$Q_t^{ex} = \left(\int_0^1 (q_t^{ex}(j))^{\frac{1}{1+\upsilon_t}} dj \right)^{1+\upsilon_t} \tag{1.27}$$

The fact that the value of the exported bundle is equal to the value of its components

$$P_t^{ex} Q_t^{ex} = \int_0^1 p_t^{ex}(j) q_t^{ex}(j) dj \tag{1.28}$$

gives the following equation for the price of the aggregate exported:

$$P_t^{ex} = \left(\int_0^1 (p_t^{ex}(j))^{-\frac{1}{\upsilon_t}} \right)^{-\upsilon_t} dj \tag{1.29}$$

As in the case of the domestic market, the intermediate producer must receive a price-changing signal to be able to reset her export price. The probability of this signal is equal to $1 - \theta_{ex}$, and the signal is completely independent of the one allowing for the reoptimization of the domestic price. The reset price is the price that maximizes the expected discounted profit from export activity:

$$\widetilde{p}_t^{ex} = \arg\max_{p_t^{ex}(j)} E_t \left[\sum_{\tau=0}^{\infty} \theta_{ex}^{\tau} \lambda_{t,t+\tau} \Pi_{t+\tau}^{ex,j} \left(p_t^{ex}(j) \right) \right] \tag{1.30}$$

where the instantaneous profit from export activity is given by the following equation:

$$\Pi_t^{ex,j} = \left(S_t p_t^{ex}(j) - MC_t\right) q_t^{ex}(j) = \left(S_t p_t^{ex}(j) - MC_t\right) \left(\frac{p_t^{ex}(j)}{P_t^{ex}}\right)^{-\frac{1+\upsilon_t}{\upsilon_t}} Q_t^{ex} \tag{1.31}$$

The first-order conditions for the optimal export reset price yield:

$$E_t \sum_{\tau=0}^{\infty} \theta_{ex}^{\tau} \lambda_{t,t+\tau} (P_{t+\tau}^{ex})^{\frac{1+\upsilon_{t+\tau}}{\upsilon_{t+\tau}}} Q_{t+\tau}^{ex} \frac{1}{\upsilon_{t+\tau}} (\widetilde{p}_t^{ex})^{-\frac{1+\upsilon_t}{\upsilon_t}-1} \times$$

$$\times \left(S_{t+\tau}\widetilde{p}_t^{ex} - (1+\upsilon_{t+\tau})MC_{t+\tau}\right) = 0 \tag{1.32}$$

Import

The importing of intermediate products is implemented by foreign companies.[4] Like domestically produced intermediate goods, all imported varieties are imperfect substitutes. The cost (in domestic currency) of importing firm j is $S_t P_t^f$, and its income is $p_t^{im}(j)$. P_t^f stands for the average cost (in foreign currency) of producing any variety abroad. Domestic prices of imported goods are also rigid due to the Calvo mechanism with price-changing probability equal to $1-\theta_{im}$. If the foreign producer is allowed to reset her price in the domestic market, she chooses the optimal level so that to maximize her expected discounted future profits (in foreign currency):

$$\widetilde{p}_t^{im} = \underset{p_t^{im}(j)}{\arg\max}\, E_t \left[\sum_{\tau=0}^{\infty} \theta_{im}^{\tau} \lambda_{t,t+\tau}^{f} \frac{\Pi_{t+\tau}^{im,j}\left(p_t^{im}(j)\right)}{S_{t+\tau}}\right] \tag{1.33}$$

where the instantaneous profit of importing firm j is given by:

$$\Pi_t^{im,j} = \left(p_t^{im}(j) - S_t P_t^f\right) q_t^{im}(j) = \left(p_t^{im}(j) - S_t P_t^f\right) \left(\frac{p_t^{im}(j)}{P_t^{im}}\right)^{-\frac{1+\upsilon_t}{\upsilon_t}} Q_t^{im} \tag{1.34}$$

where foreign importers are assumed to be risk-neutral, so they discount their profits at the international risk-free rate:

$$\lambda_{t,t+\tau}^{f} = \prod_{j=t}^{t+\tau-1} \left(1 + i_j^f\right)^{-1} \tag{1.35}$$

where i_t^f is a foreign risk-free rate that is defined exogenously.

[4] Postulating this, we follow Dam and Linaa (2005). Kollmann (2001) implicitly assumes that domestic firms are engaged both in importing and exporting activities.

The first-order conditions for the problem facing the foreign importers result in the following equation for the optimal import price:

$$E_t \sum_{\tau=0}^{\infty} \theta_{im}^{\tau} \lambda_{t,t+\tau}^{f} \frac{1}{S_{t+\tau} \upsilon_{t+\tau}} (P_{t+\tau}^{im})^{\frac{1+\upsilon_{t+\tau}}{\upsilon_{t+\tau}}} Q_{t+\tau}^{im} \widetilde{p}_t^{im}(j)^{-\frac{1+\upsilon_{t+\tau}}{\upsilon_{t+\tau}}-1} \times$$
$$\times \left(\widetilde{p}_t^{im}(j) - (1+\upsilon_{t+\tau}) S_{t+\tau} P_{t+\tau}^{f} \right) = 0 \quad (1.36)$$

As cost functions are identical for any firm in the intermediate goods and foreign sectors, all producers that have the opportunity to reoptimize their prices at time t, set them at the same level ($\widetilde{p}_t^d(j) = \widetilde{p}_t^d$, $\widetilde{p}_t^{ex}(j) = \widetilde{p}_t^{ex}$ and $\widetilde{p}_t^{im}(j) = \widetilde{p}_t^{im}$ for all j). Therefore, the price indices of domestic, export and import aggregates are given by the following equations:

$$\left(P_t^d\right)^{-\frac{1}{\upsilon}} = \theta_d \left(P_{t-1}^d\right)^{-\frac{1}{\upsilon}} + (1-\theta_d) \left(\widetilde{p}_t^d\right)^{-\frac{1}{\upsilon}} \quad (1.37)$$

$$\left(P_t^{ex}\right)^{-\frac{1}{\upsilon}} = \theta_{ex} \left(P_{t-1}^{ex}\right)^{-\frac{1}{\upsilon}} + (1-\theta_{ex}) \left(\widetilde{p}_t^{ex}\right)^{-\frac{1}{\upsilon}} \quad (1.38)$$

$$\left(P_t^{im}\right)^{-\frac{1}{\upsilon}} = \theta_{im} \left(P_{t-1}^{im}\right)^{-\frac{1}{\upsilon}} + (1-\theta_{im}) \left(\widetilde{p}_t^{im}\right)^{-\frac{1}{\upsilon}} \quad (1.39)$$

1.2.3 Households

The population is assumed to consist of a continuum of households of unity measure. Any representative household maximizes its expected discounted utility over an infinite horizon subject to its budget constraints. The utility function is increasing in consumption and decreasing in labor efforts. Only final good can be consumed.

We follow many other papers (Erceg, Henderson, and Levin (2000), Gali (2008)) in assuming that labor services of different households are imperfect substitutes, as indicated above. Every household holds monopoly power in the market over its variety of labor and acts as a wage-setter. A wage-setting process is also rigid à Calvo with the probability of a wage-changing signal equal to $1 - \theta_w$.

Each period, a representative household makes its consumption and portfolio choices. A household can own domestic and foreign bonds[5] as well as capital. If a household receives a wage-changing signal, it also makes a decision about a new reset price. A household faces only one kind of uncertainty – when it will be allowed to change its wage for the next

[5] We assume incomplete financial markets.

time – and this shock is idiosyncratic. Therefore, different households can work different amounts of time and have different incomes (Christiano et al. (2005)). But, as was shown in Woodford (1996) and Erceg et al. (2000), we can consider households to be homogenous with respect to the amount of consumption and wealth allocation among different types of bonds and capital owing to state-contingent assets. It allows us to drop a household index h for consumption in the utility function.

A household h maximizes its expected discounted utility (subject to the budget constraint to be specified below):

$$V_0(h) = \max E_0 \sum_{t=0}^{\infty} \beta^t U\left(C_t, l_t(h)\right) \tag{1.40}$$

where C_t represents consumption, $l_t(h)$ is the labor services supplied by household h, and β is a subjective discount factor. As indicated above, the household manages three kinds of assets: domestic bonds, foreign bonds and capital stock. In addition to interest on bonds and capital, a household receives labor income, dividends from non-competitive intermediate firms, and revenues from commodity exports.

The capital accumulation equation can be written as:

$$K_{t+1} = (1-\delta)K_t + I_t - \chi\left(K_{t+1} - K_t\right) \tag{1.41}$$

where I_t is investment, and δ is the depreciation ratio. The last term in (1.41) stands for the capital adjustment cost, and the function χ is defined as follows:

$$\chi\left(K_{t+1}, K_t\right) = \frac{\Phi}{2}\frac{\left(K_{t+1} - K_t\right)^2}{K_t} \tag{1.42}$$

We follow Smets and Wouters (2003) in defining the preferences, which are assumed to be described by an additively separable instantaneous utility function with CRRA form:

$$U\left(C_t, l_t(h)\right) = \epsilon_b \left(\frac{\left(C_t - \nu\widetilde{C}_{t-1}\right)^{1-\sigma_1}}{1-\sigma_1} - \epsilon_l \frac{l(h)^{1+\sigma_2}}{1+\sigma_2} \right) \tag{1.43}$$

letting $\widetilde{C}_{t-1}$ be external habits in consumption (Abel (1990)) and letting ν be a positive parameter of force of habits. The budget constraint of household h in period t is represented by the following equation:

$$P_t(C_t + I_t(h)) + D_t(h) + S_t D_t^*(h) =$$
$$\int_0^1 w_t(h) l_t(h,j)dj + D_{t-1}(h)\left(1 + i_{t-1}\right) + S_t D_{t-1}^*(h)\left(1 + i_{t-1}^*\right) +$$
$$R_t^K K_t(h) + \Pi_t^d(h) + \Pi_t^{ex}(h) + S_t O_t(h) \quad (1.44)$$

The commodity production is assumed to be constant and normalized to unity, so all the fluctuations of commodity export revenues are due to changes of the commodity price (denoted by O_t in this paper). D_t^* denotes foreign bonds (credit from the foreign sector if D_t^* is negative), i_t is the nominal domestic interest rate, and i_t^* is the nominal foreign interest rate (including the risk premium). The financial markets are assumed to be imperfect, and the imperfections create a deviation of nominal interest rate on foreign bonds from the international risk-free rate i_t^f. This deviation can be interpreted as a risk premium:

$$1 + i_t^* = \rho\left(1 + i_t^f\right) \quad (1.45)$$

Like Lindé, Nessen, and Soderstrom (2009) and Curdia and Finocchiaro (2005), we assume that this risk premium can be specified by a decreasing function of net foreign assets of the economy. However, unlike the cited papers, we modify the function of risk premium and normalize net foreign assets to the total export (including commodity export income) in steady state:

$$\rho_t = \exp\left(-\omega\left(\frac{\bar{P}^f D_t^*}{\bar{P}^{ex}\bar{Q}^{ex} + \bar{O}}\right) + \epsilon_t^\rho\right) \quad (1.46)$$

where ϵ_t^ρ is a stochastic shock of the risk premium, ω is a normalizing constant, and barred variables here and below denote steady-state values of the corresponding variables without bars. Therefore, if the amount of debt of domestic households increases, the interest rate (with premium) increases as well. The technical reason for including the endogenous risk premium is that it guarantees the existence of stationary equilibrium (Schmitt-Grohe and Uribe (2003)).

During each period, a representative household maximizes its expected discounted utility (1.40) subject to the sequence of dynamic constraints: (1.44) and (1.41).

The first-order conditions for this problem yield the following equations:

$$U'_C = P_t \mu_t \tag{1.47}$$

$$\beta E_t \mu_{t+1}(1+i_t) = \mu_t \tag{1.48}$$

$$\beta E_t \mu_{t+1} S_{t+1}(1+i^*_t) = \mu_t S_t \tag{1.49}$$

$$\beta E_t \mu_{t+1} R^K_{t+1} + \beta E_t P_{t+1}\mu_{t+1}(\left(1-\delta-\chi'_{2,t+1}\right) = \mu_t(1+\chi'_{1,t})P_t \tag{1.50}$$

where μ_t is the Lagrange multiplier on the budget constraint. As indicated above, the household decides on consumption, investment, and portfolio distribution every period, but it chooses an optimal wage only on occasion when a wage-changing signal occurs. To derive the optimal reset wage for the firm reoptimizing in period t, we reproduce the relevant parts of the maximization problem written above. We take into account the probability that a new wage-changing signal does not come until $t+s$ is θ^s_w. In periods of wage resetting, the household maximizes the expected discounted utility:

$$V^w_t(h) = \max E_t \sum_{\tau=0}^{\infty} (\beta\theta_w)^\tau U\left(C_{t+\tau|t}, l_{t+\tau|t}(h)\right) \tag{1.51}$$

subject to the sequence of labor demand and budget constraints:

$$l_{t+\tau|t}(h,j) = L_{t+\tau|t}(j)\left(\frac{\widetilde{w}_t(h)}{W_{t+\tau|t}}\right)^{-\frac{1+\gamma}{\gamma}} \tag{1.52}$$

$$\begin{aligned} P_{t+\tau|t}(C_{t+\tau|t} + I_{t+\tau|t}(h)) + D_{t+\tau|t}(h) + S_{t+\tau|t}D^*_{t+\tau|t}(h) = \\ \int_0^1 w_{t+\tau|t}(h) l_{t+\tau|t}(h,j)dj + D_{t+\tau-1|t}(h)\left(1+i_{t+\tau-1|t}\right) + \\ S_{t+\tau|t}D^*_{t+\tau-1|t}(h)\left(1+i^*_{t+\tau-1|t}\right) + R^K_{t+\tau|t}K_{t+\tau|t}(h) + \Pi^d_{t+\tau|t}(h) + \\ \Pi^{ex}_{t+\tau|t}(h) + S_{t+\tau|t}O_{t+\tau|t}(h) \end{aligned} \tag{1.53}$$

The first-order conditions for this problem result in the following equation for the reset wage:

$$\widetilde{w}_t(h)^{\frac{1+\gamma}{\gamma}\sigma_2+1} = (1+\gamma)\frac{E_t\sum_{\tau=0}^{\infty}\beta^\tau\theta^\tau_w \epsilon^b_{t+\tau}\epsilon^l_{t+\tau}L^{1+\sigma_2}_{t+\tau}W^{\frac{1+\gamma}{\gamma}(1+\sigma_2)}_{t+\tau}}{E_t\sum_{\tau=0}^{\infty}\beta^\tau\theta^\tau_w\mu_{t+\tau}L_{t+\tau}W^{\frac{1+\gamma}{\gamma}}_{t+\tau}} \tag{1.54}$$

where μ_t is the Lagrange multiplier associated with the budget constraint as given above. As the function of optimal wage does not depend on h, all households that have the opportunity to reoptimize their wages at time t, set them at the same level ($\widetilde{w}_t(h) = \widetilde{w}_t$). With the premise that, in each period, the ratio of households adjusting their wage is equal to $1 - \theta_w$, the law of motion for aggregate wages can be derived as:

$$(W_t)^{-\frac{1}{\gamma}} = \theta_w \left(W_{t-1}\right)^{-\frac{1}{\gamma}} + (1 - \theta_w) \left(\widetilde{w}_t\right)^{-\frac{1}{\gamma}} \tag{1.55}$$

1.2.4 Central bank

Because the goal of this paper is to estimate a DSGE model for the Russian economy, it is very important to use a monetary rule that actually describes the strategy of the Bank of Russia. However, at the moment, there is no scientific consensus regarding the monetary policy rule of the Bank of Russia. In the economic literature, the absence of a common opinion is indicated by the existence of different points of view regarding the best way to model the central bank's activity. For example, Vdovichenko and Voronina (2006) show that from 1999 to 2003, the Bank of Russia regulated the money supply, while the use of monetary instruments was limited by interventions on exchange markets and the sterilization of excess liquidity with deposit operations. The authors claim that, unlike most central banks in developed countries, the discount rate in Russia plays a minor role. Hence they opt for the money supply rule. On the contrary, Benedictow, Fjaertoft, and Lofsnaes (2013) estimate an econometric model of the Russian economy based on data from 1995 to 2008. They suppose that monetary policy follows a simple Taylor rule, and the interest rate is set in response to unemployment and inflation changes. The authors claim that this kind of rule fits the data well even though the assumptions in the basis of the rule are hardly relevant to the Russian economy. In line with Benedictow et al. (2013), Taro (2010) successfully estimate a non-linear interest rate rule on data from 1997 to 2007 under the assumption that the reaction of the central bank to an output gap and inflation is asymmetric. Finally, Yudaeva, Ivanova, and Kamenskikh (2010) aim to determine a monetary policy target for the Bank of Russia. They show that a forward-looking Taylor rule, as well as a money supply rule, can adequately describe Russian data from 2003 to 2010. Their results demonstrate that the central bank sets its instrument in response to the expected movements of inflation, output, and exchange rate, and uses interest rate smoothing. The authors do not opt for either of these rules. However, the fact that there are econometric papers that show that an interest rate rule can describe monetary policy in Russia allows us to use this kind of rule in our structural model. We therefore assume that monetary policy follows a modified Taylor rule with interest rate smoothing:

$$1 + i_t^d = (1 + i_{t-1}^d)^{z_1}(1 + \bar{i}^d)^{1-z_1} \left(\frac{\pi_t}{\bar{\pi}}\right)^{z_2(1-z_1)} \left(\frac{Y_t}{\bar{Y}}\right)^{z_3(1-z_1)} \epsilon_t^m \tag{1.56}$$

where π_t denote inflation, and z_1, z_2, z_3 are positive parameters with $z_1 \leq 1$.

1.2.5 Market clearing conditions and exogenous processes

During each period, an equilibrium in goods and financial markets must be maintained and a balance-of-payment identity must hold. Domestically produced intermediate goods are consumed within the economy or exported:

$$Y_t = Q_t^d + Q_t^{ex} \tag{1.57}$$

The final good is divided among consumption and investment:

$$Q_t = C_t + I_t \tag{1.58}$$

The balance-of-payment identity is derived from the household's budget constraint (1.44) and the equation of final good allocation (1.57). The balance-of-payment identity takes the form of:

$$P_t^{ex} Q_t^{ex} + O_t - \frac{1}{S_t} P_t^{im} Q_t^{im} - D_t^* + \left(1 + i_{t-1}^f\right) D_{t-1}^* = 0 \tag{1.59}$$

This equation implies that the exchange rate is floating. We are aware of the fact that this is not the case in Russia, but we think that complicating of the model may not make the estimation more accurate. We assume that all exogenous processes, except mark-up and monetary policy shocks, are given by AR(1) and the mark-up shock and monetary policy shocks are i.i.d processes:

$$\log A_t = \rho_a \log A_{t-1} + (1-\rho_a) \log \bar{A} + \varepsilon_t^a \tag{1.60}$$

$$\log O_t = \rho_o \log O_{t-1} + (1-\rho_o) \log \bar{O} + \varepsilon_t^o \tag{1.61}$$

$$\log Y_t^f = \rho_{y^f} \log Y_{t-1}^f + (1-\rho_{y^f}) \log \bar{Y}^f + \varepsilon_t^y \tag{1.62}$$

$$\log \pi_t^f = \rho_{\pi^f} \log \pi_{t-1}^f + (1-\rho_{\pi^f}) \log \bar{\pi}^f + \varepsilon_t^\pi \tag{1.63}$$

$$\log (i_t^f + 1) = \rho_{i^f} \log (1 + i_{t-1}^f) + (1-\rho_{i^f}) \log (1 + \bar{i}^f) + \varepsilon_t^i \tag{1.64}$$

$$\log \epsilon_t^b = \rho_b \log \epsilon_{t-1}^b + (1-\rho_b) \log \bar{\epsilon}^b + \varepsilon_t^b \tag{1.65}$$

$$\log \epsilon_t^l = \rho_l \log \epsilon_{t-1}^l + (1-\rho_l) \log \bar{\epsilon}^l + \varepsilon_t^l \tag{1.66}$$

$$\log \epsilon_t^\rho = \rho_z \log \epsilon_{t-1}^\rho + (1-\rho_\rho) \log \bar{\epsilon}^\rho + \varepsilon_t^\rho \tag{1.67}$$

$$\log \upsilon_t = \log \bar{\upsilon} + \varepsilon_t^\upsilon \tag{1.68}$$

$$\log \epsilon_t^m = \varepsilon_t^z \tag{1.69}$$

Finally, our measure of real GDP in the model is:

$$GDP_t = Q_t + \frac{S_t P_t^{ex} Q_t^{ex} + S_t O_t - P_t^{im} Q_t^{im}}{P_t}. \tag{1.70}$$

1.3 Estimation

To find a solution for the model, we normalize all the nominal variables to national or foreign price levels (see Appendix A) and log-linearize the non-linear system around a non-stochastic steady state. We assume that in a steady state the current account is equal to zero; we also assume that $\eta = 1$. These assumptions are sufficient to derive an analytical solution for all the variables in a steady-state. We present the steady-state derivation in Appendix B and the final log-linearized model in Appendix C. We solve the model in Dynare and estimate it using Bayesian techniques. We think that calibration is unsuitable in our case because of a lack of microeconomic and macroeconomic papers that could have served as references for assigning values to hyperparameters.

1.3.1 Solution and data

As in many other DSGE papers, the estimation is done using Bayesian techniques. After choosing the prior, we combine it with the sample information described by the likelihood function and we receive the posterior distribution, by employing Bayes theorem. For the sake of simplicity, we characterize the posterior distribution by its mode, median, and variance. The posterior distribution is estimated in two steps. First, the posterior mode and approximate covariance matrix are calculated. The covariance matrix is computed numerically as the inverted (negative) Hessian at the posterior mode. Thereafter, the posterior distribution of model parameters is generated with a random-walk Metropolis–Hastings algorithm.

There are ten shocks in the model: technology shock, commodity export revenues shock, monetary policy shock, mark-up shock, preference shock, labor supply shock, foreign interest rate shock, foreign prices shock, foreign output shock, and risk premium shock. For our estimation, we use nine time series. This guarantees the absence of stochastic singularity without resorting to measurement errors. Thus, we implicitly assume that all the observed volatility is caused by structural shocks. The variables that we consider to be observed for estimation are consumption, domestic inflation, domestic interest rate, real wages, the real exchange rate, oil revenues, foreign inflation, foreign interest rate, and foreign output.

The source for most of the data is the International Financial Statistics database. Other sources will be indicated below. All the series are quarterly, starting in the third quarter of 1999 and ending in the third quarter of 2011 [6]. We take into account the fact that the series are quite short, but we intentionally avoided using the earlier data on account of the severe financial crisis of 1998. By the third quarter of 1999, the impact of the financial crisis of 1998 on the Russian economy was reduced substantially. This allows us to consider our sample period (at least before 2008) as relatively homogenous both in terms of policy and hitting shocks. We are aware of the fact that the Bank of Russia changed its monetary policy after the financial crisis of 2008, but we do not restrict the sample intentionally to the end of 2008 to avoid making our time series even shorter. To convert nominal variables (consumption, output) to real terms, we use the GDP deflator. The series were seasonally adjusted with Census X12.

As an observable series for consumption, we use nominal private final consumption expenditures per capita. After seasonal adjustment, we take the logarithm of the series

[6] The paper was started in 2012 and accepted for publication in 2013.

and detrend it linearly. The series of linearly detrended producer price inflation stands for an observable series of domestic inflation (π_t^d). The interest rate is assumed to correspond to the money market rate. The quarterly values were calculated by dividing the annual (detrended) interest rate (in percentage points) by 400. The series for wages was taken from the Rosstat database. The series is already seasonally adjusted, so we make no additional adjustment; we just take the logarithm and detrend the series linearly. For the real exchange rate series ($\mathcal{E}_t$)[7] we take the weighted average of EUR/RUR and USD/RUR exchange rate series. The weights are 0.45 and 0.55, respectively. These are the same weights that the Bank of Russia has used for calculating the currency basket (the operational benchmark for the exchange rate policy) since February 2007 [8]. The series of real dollar and euro exchange rates were calculated on the basis of consumer price indices. Finally, we take the logarithm and detrend linearly the series of the exchange rate.

Next, we take per capita revenues from the export of crude oil, oil products, and natural gas to stand for the observable variable of commodity export. The data source on commodity exports is Balance of Payments statistics provided by the Bank of Russia. The series is expressed in terms of the bi-currency basket; we take the logarithm of the series and detrend it linearly.

All foreign variables are also expressed in terms of the bi-currency basket. The CPI inflation series for both the U.S. and the euro area are combined to stand for the foreign inflation variable in the model. We use money market rates for the euro area and the US (federal funds rate) to calculate the series for the foreign interest rate. The annual series (in percentage points) are divided by 400. To calculate the series of observable output for the foreign economy, we use weighted per-capita GDP values. We seasonally adjust the

[7] Increase of the variable means real depreciation of the rouble.

[8] The policy of using the bi-currency basket as an operational target started in February 2005. Before February 2005 the exchange rate vis-à-vis the US dollar was an operational target. During 2005–2007, the shares of the dollar and the euro in the basket were revised five times, and the share of the euro increased. In this paper, we intentionally ignore these revisions to avoid artificial jumps in the series denominated in the foreign currency. Moreover, in estimation of DSGE, the common practice is to use the effective exchange rate. According to our preliminary estimates, the average share of exports and imports of the euro area and Switzerland among 15 major trading partners of Russia in 1999–2011 was 45.2% (our calculations based on IMF data).

series, take logarithms, and detrend it linearly. In our calculations, we consider 2005 as a base year, which does not affect the calculations, as for all series (except interest rates and inflation), we take the logarithm of the series and detrend them.

1.3.2 Priors

When choosing the prior distributions, we follow the common practice in the literature. We fix the same subset of parameters that is usually fixed in similar studies. In a Bayesian sense, this means that we assign zero variance of prior distribution, so we set the discount factor β at 0.99 and the depreciation rate at 0.025. We fix ψ (the capital elasticity of the production function) at 0.33, which reflects the scientific consensus that the ratio of labor to overall income is about two-thirds. As P_{ex} and P_{im} are determined by the same shock as P_d, we assume that θ_{ex} and θ_{im} are equal to θ_d, which is estimated. We tried to estimate the ratio of domestic goods in consumption (α_d), but the estimated value was too low and the whole convergence deteriorated, so we fix α_d at 0.74 according to calculations made in our previous studies (Malakhovskaya (2013)). Following Dam and Linaa (2005), we also fix the net wage mark-up and steady-state value of the net price mark-up process at 0.2. Our system includes the value of the steady-state of oil revenues, which is not known. To overcome the problem, we rewrite the system in terms of $\tilde{o} = \frac{\bar{O}}{\bar{P}^{ex}\bar{Q}^{ex}}$ and calibrate $\tilde{o}$ as the mean value of the ratio of commodity exports (crude oil, oil products, and natural gas) to non-commodity exports (all exports besides revenues from crude oil, oil products, and natural gas) over the sample period.[9]

We estimate 28 parameters in total, which are the parameters of preference, production function, and capital adjustment cost, as well as autocorrelation coefficients and standard errors that determine structural shocks. While choosing the prior distributions, we follow common rules: we assume beta distribution for all the parameters that can take only values between 0 and 1, we assume gamma distribution for all preference parameters, normal distribution for parameters of the monetary policy rule, and inverted gamma distribution for standard errors of structural shocks. When choosing moments of prior distributions, we follow Smets and Wouters (2003), Smets and Wouters (2007), Dam and Linaa (2005)

[9]For the chosen value $\tilde{o}$, the ratio of commodity revenues to GDP in the model is equal to 14.7%. The actual mean value over the sample is 17.6%.

whenever it is possible. The fact that the moments of posterior distribution are different from those in cited papers means that the estimates are primarily determined by data and not by prior distributions. For other variables, the mean values of prior distributions were chosen to be consistent with our econometric estimates.

We follow existing studies in assuming that θ_d and θ_w have a beta distribution with a mean set at 0.75 and a standard deviation at 0.1 (for example, Smets and Wouters (2003)). This implies that prices and wages are reset about once a year.

We assume that the mean values of prior distributions for utility parameters (σ_1 and σ_2) are equal to 1 and 2, respectively, following Smets and Wouters (2003). We also assume that the habit persistence parameter fluctuates around 0.6, with a standard deviation equal to 0.1. We intentionally choose a smaller mean value than in Smets and Wouters (2003) and Smets and Wouters (2007) to make the prior distribution more symmetric because of a lack of any previous estimates of this parameter in the Russian data. Our prior distribution for the capital adjustment cost parameter corresponds to that in Dam and Linaa (2005), but contrary to Dam and Linaa (2005), we estimate the capital mobility parameter, and we set the mean value of its prior distribution at 0.002 following Lane and Milesi-Ferretti (2001). The parameters of the monetary policy rule are assumed to be normally distributed. The mean values of prior distributions generally correspond to a simple Taylor rule and are the same as in Smets and Wouters (2007), for instance, but we assume greater standard deviations than in existing papers because it allows us to admit a wider range of possible parameters for the rule. To determine the mean values of prior distributions for autocorrelation parameters for the commodity export revenues process (ρ_o) and all exogenous processes describing the foreign economy ($\rho_{y^f}, \rho_{i^f}, \rho_{y^f}$), we use regressions on our data. For these four parameters, we choose small standard deviations to make their distributions tight. As for all remaining autocorrelation coefficients, we assume that they have a beta distribution with a mean value set at 0.5 and a standard deviation set at 0.2, in accordance with Smets and Wouters (2007). All standard errors of exogenous process are assumed to follow an inverted gamma distribution. We choose the same mean value for all distributions except one. For σ_{i^f}, we take a smaller value because of the convergence problem.

1.3.3 Estimation results

We summarize our assumptions about the prior distributions and present our estimation results in Table 1.

First of all, we present means and standard deviations of prior distributions. Then we show the mode and standard error of posterior distributions, which are estimated by a numerical minimization method. The standard error is calculated on the basis of Hessian estimated at the mode of distribution. Finally, we present the median and 80% interval for each parameter. These values were estimated with the MCMC algorithm. The Metropolis-Hastings algorithm was implemented with 400,000 iterations with two chains. But the convergence was achieved earlier, which can be confirmed by Brooks and Gelman's procedure. [10]

All the estimates are significantly different from zero. For all prior and posterior distributions, see Appendix C. They confirm that the convergence is good. For all autocorrelation coefficients for structural shocks except two (ρ_b and ρ_{π_f}), the mode values are higher than 0.7. This validates the hypothesis about the high persistence of structural shocks.

In addition, our estimations of nominal rigidity parameters (θ_d) and (θ_w) do not contradict economic logic (about 0.5 and 0.4, respectively). This means that prices and wages are not very rigid, with contracts lasting about 5 months for wages and 8 months for prices. It is noteworthy that our estimates differ from the estimates of nominal rigidity parameters in other papers, where the level of nominal rigidity turned out to be unreasonably high. For example, in the paper by Dam and Linaa (2005), which is very close to ours with regard to the theoretical model, the estimate of the nominal rigidity parameter is 0.94, which means that contracts are not reset for about four years. The fact that our estimates of the nominal rigidity parameter are not too big allows us not to resort to inflation indexation in the Calvo mechanism, as in Christiano et al. (2005). Besides, in the paper by Dam and Linaa (2005), the authors received a very high value of mark-up volatility. Our estimate of this parameter is completely reasonable.

All the remaining parameters also take reasonable values. For example, the habit formation parameter is estimated to be 0.66. This value is higher than estimates in

[10] The univariate MCMC diagnostics can be sent by the authors upon request.

Table 1. Prior and posterior distribution of parameters.

Parameter	Prior distribution			Posterior estimate		Posterior distribution		
	Type	Mean	St. dev.	Mode	St. error	10%	Median	90%
θ_d	beta	0.75	0.1	0.507	0.084	0.424	0.546	0.687
θ_w	beta	0.75	0.1	0.398	0.065	0.359	0.442	0.523
σ_1	gamma	1	0.3	1.015	0.25	0.819	1.126	1.521
σ_2	gamma	2.0	0.6	1.74	0.513	1.275	1.884	2.68
ϕ	gamma	15	4	10.57	4.44	6.87	12.05	18.23
ν	beta	0.6	0.1	0.661	0.086	0.549	0.662	0.757
ω	normal	0.002	0.001	0.0033	$8.6 \cdot 10^{-4}$	0.0023	0.0034	0.0045
z_1	beta	0.8	0.1	0.861	0.029	0.823	0.862	0.894
z_2	normal	1.5	0.3	1.597	0.246	1.295	1.607	1.929
z_3	normal	0.12	0.075	0.103	0.051	0.047	0.11	0.187
ρ_{y^f}	beta	0.94	0.01	0.943	0.01	0.929	0.942	0.954
ρ_{π^f}	beta	0.28	0.01	0.28	0.01	0.268	0.28	0.293
ρ_b	beta	0.5	0.2	0.351	0.141	0.202	0.367	0.537
ρ_l	beta	0.5	0.2	0.888	0.078	0.68	0.853	0.933
ρ_a	beta	0.5	0.2	0.862	0.069	0.694	0.829	0.918
ρ_o	beta	0.75	0.05	0.787	0.043	0.728	0.785	0.837
ρ_{i^f}	beta	0.98	0.01	0.972	0.013	0.95	0.969	0.983
ρ_{rp}	beta	0.5	0.2	0.741	0.065	0.623	0.719	0.799
σ_{ea}	inv. gam.	0.05	Inf	0.032	0.011	0.025	0.04	0.085
σ_{i^f}	inv. gam.	0.02	Inf	0.002	$3.8 \cdot 10^{-4}$	0.0024	0.0026	0.0029
σ_{π^f}	inv. gam.	0.05	Inf	0.007	$6.8 \cdot 10^{-4}$	0.006	0.007	0.008
σ_{y^f}	inv. gam.	0.05	Inf	0.009	$8.4 \cdot 10^{-4}$	0.008	0.009	0.01
σ_{eb}	inv. gam.	0.05	Inf	0.081	0.02	0.067	0.091	0.126
σ_{el}	inv. gam.	0.05	Inf	0.257	0.078	0.236	0.358	0.539
σ_{ez}	inv. gam.	0.05	Inf	0.012	0.002	0.01	0.012	0.014
σ_{eo}	inv. gam.	0.05	Inf	0.13	0.013	0.117	0.132	0.15
σ_{er}	inv. gam.	0.05	Inf	0.017	0.004	0.015	0.019	0.025
σ_{ev}	inv. gam.	0.05	Inf	0.016	0.04	0.013	0.01	0.03

Smets and Wouters (2003) for the euro area (0.541) and in (Dam and Linaa (2005)) for Denmark (0.433). This fact can be interpreted as a higher inertia of consumption in Russia. The parameters of preferences (σ_1 and σ_2) also took the plausible values of 1.01 and 1.73, respectively. This means that the labor supply elasticity is about 0.6 and the intertemporal elasticity of substitution is unity. It is worth noting that our estimate of labor supply elasticity is less than values usually used to calibrate macroeconomic models, but it corresponds well to microeconometric estimates of this parameter (Peterman (2012)). All parameter estimates of the monetary policy rule are also in line with economic logic.

1.3.4 Impulse response analysis and historical decomposition

Impulse response analysis

After estimating the model, we analyzed its properties with impulse response functions.

In Figure 1, we present the effects of a positive shock of commodity exports on the dynamics of the main macroeconomic variables.

The increase in households' income implies an increase in households' consumption and their demand in goods market. This is followed by an increase in labor demand, investments, capital, wages, the rate of return on capital, GDP, and domestic output (without oil). The rise in commodity export revenues results in a real appreciation of the exchange rate, encouraging non-commodity imports and discouraging non-commodity exports. In quantitative terms (in percentage points), a positive shock of commodity export revenues has the strongest influence on GDP, real exchange rate, investments, exports and imports. Domestic production changes positively, though only to a small extent, but the effect persists.

Historical decomposition

In this section, we investigate what the driving forces of the main macroeconomic variables in Russia are. The model can describe which shocks dominate the dynamics of all observed variables. Figures 2–4 show the historical contribution of all shocks to some variables of interest with columns of different patterns. In Figure 2, the historical decomposition of the detrended logarithm of consumption over the sample period is

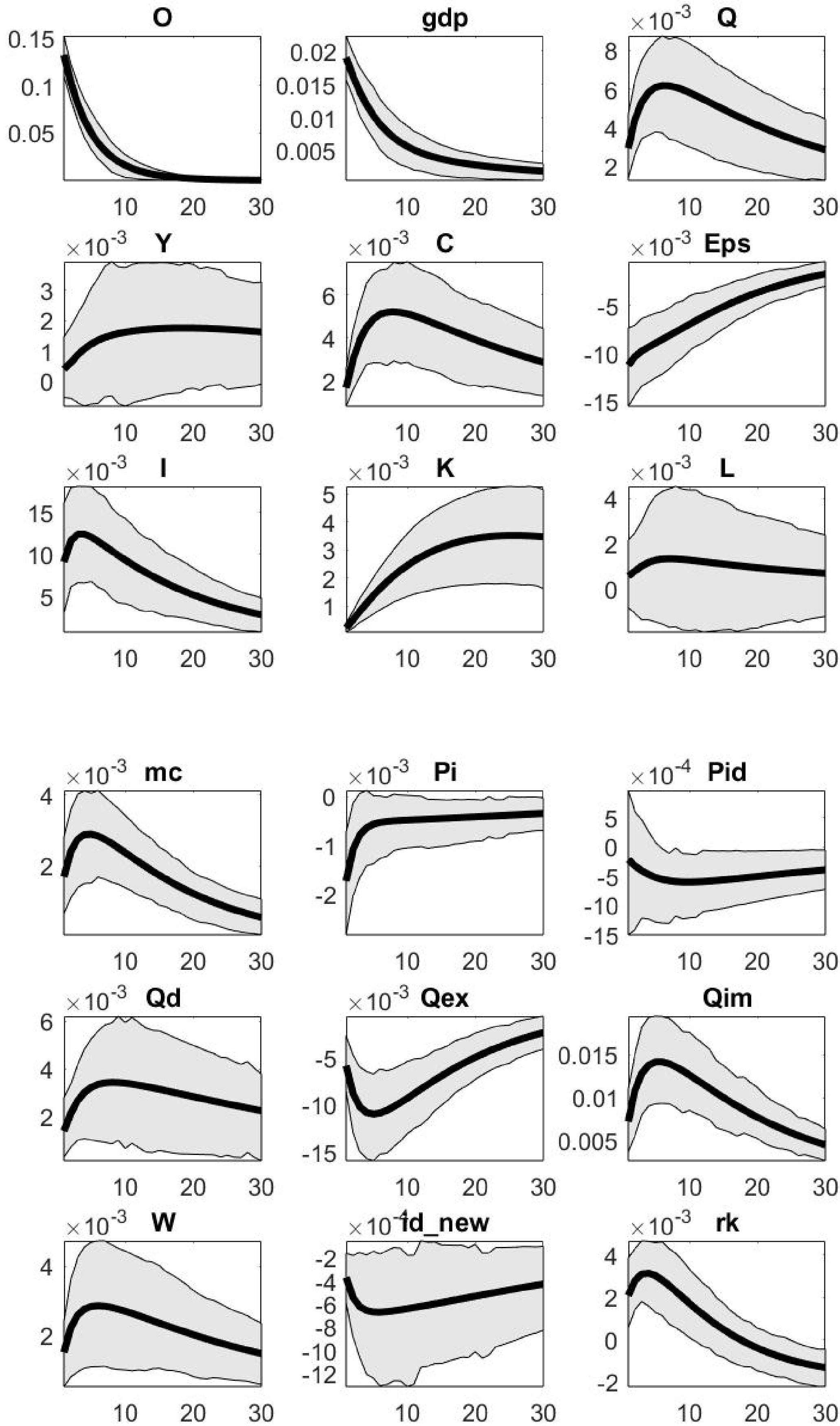


Figure 1.1: Oil price shock effect, impulse response functions with 90% HPDI.

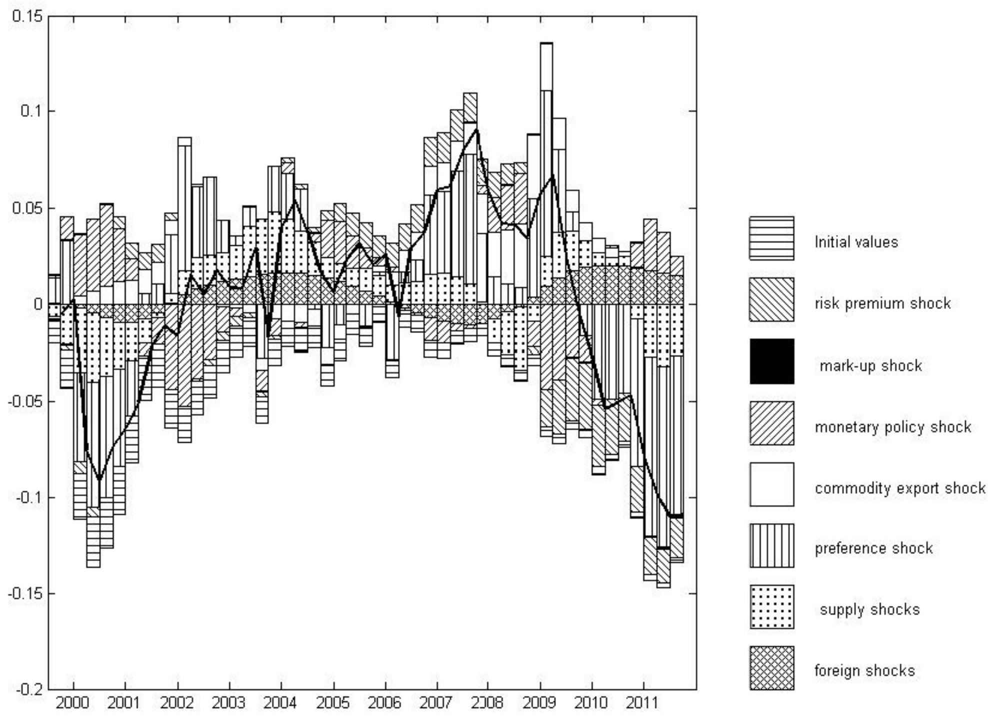


Figure 1.2: Historical decomposition of consumption

presented.

It is noteworthy that, despite the fact that the level of openness of the Russian economy is rather high (primarily due to oil exports), the dynamics of consumption is explained, first of all, by domestic shocks. The shock of preferences and technology shocks are the most influential for consumption dynamics. As expected, the commodity export revenues shock is relatively important. This shock contributed to a great extent to consumption growth during the four years before the financial crisis of 2009.[11]

In Figure 3, we present the historical decomposition of the logarithm of the detrended real exchange rate. The figure shows that the commodity shock contributes more strongly to the RER error variance than to the consumption error variance. We pay attention to the fact that during the four years before the crisis, the commodity export shock contributed to real appreciation of the exchange rate. The figure also shows that the

[11]The financial crisis hit Russia later than the US and Europe.

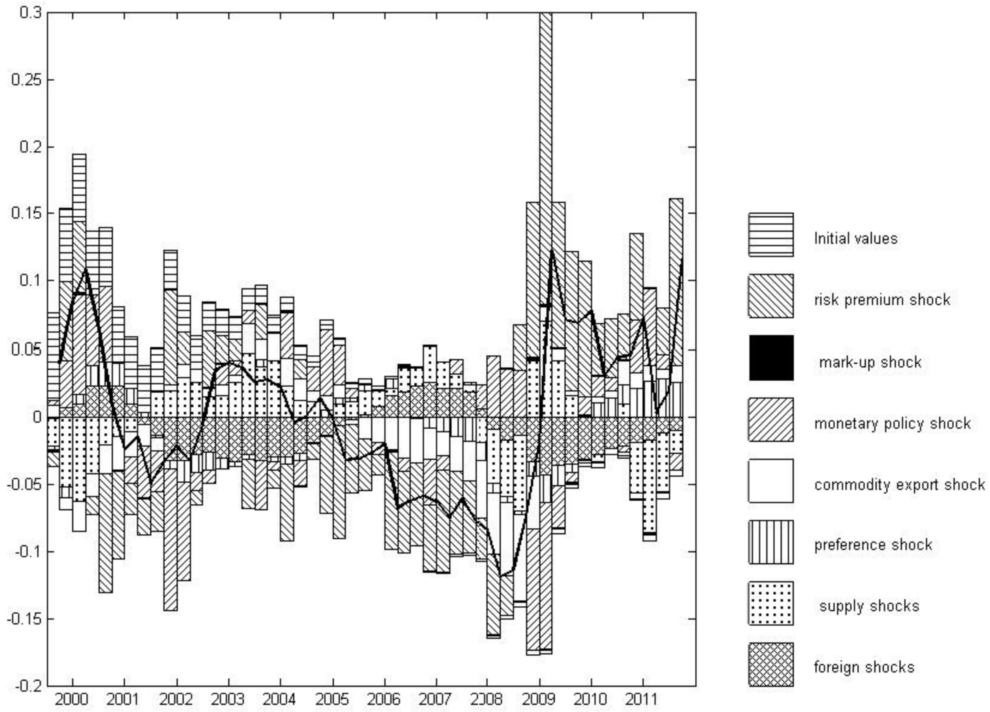


Figure 1.3: Historical decomposition of real exchange rate

abrupt depreciation of the rouble in the first quarter of 2009 was caused by a sharp increase in risk premium, followed, consequently, by a small negative effect of commodity export shock. The monetary policy of the central bank probably helped to avoid even greater depreciation than could have taken place.

In Figure 4, the historical decomposition of a simulated series of GDP can be found. We simulate the series because we do not have it among our observable variables. Figure 4 shows that the commodity shock contributes much to GDP dynamics over the sample period. It is noteworthy that the commodity export shock explains the output growth before the financial crisis. The figure also shows that the output decrease in 2009 was caused by the joint pressure of negative commodity export shock and restrictive monetary policy (interest rate increase).

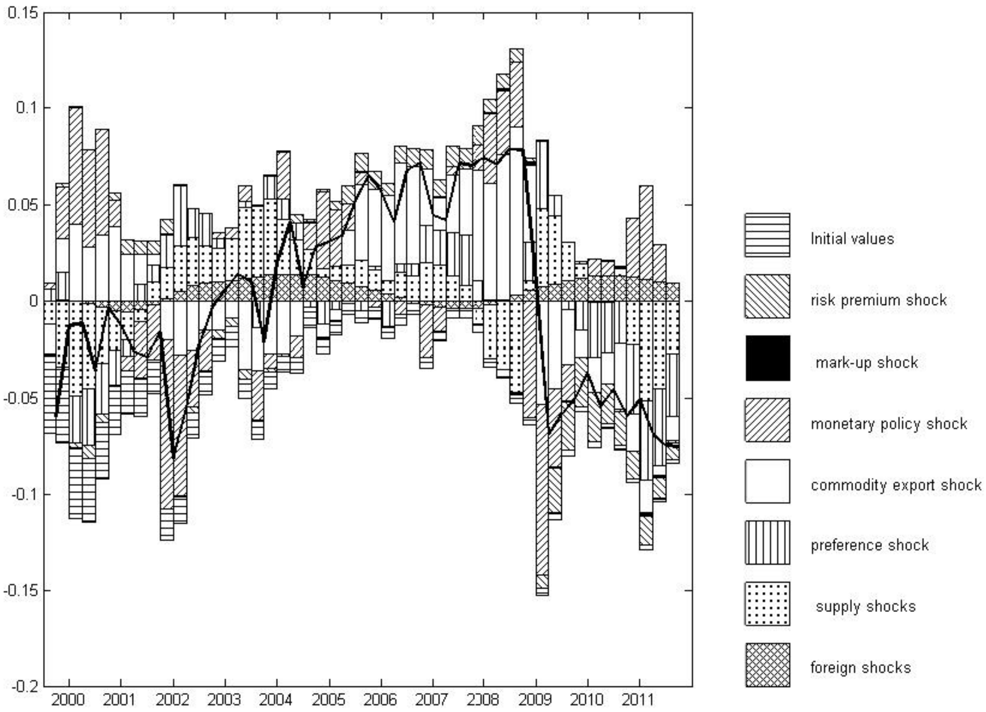


Figure 1.4: Historical decomposition of GDP (simulated series)

Forecast error variance decomposition

Table 2 shows a variance decomposition of forecast error at various horizons: in the short run (1 year), medium run (3 years), and long run (20 years). This allows us to come to several conclusions. First of all, technological and labor supply innovations explain a large part of all measures of output in the model, including GDP, output of final goods (without oil), and output of intermediate goods and at all horizons. Supply shocks account for 68% of the error variance of intermediate goods output in the short run and for more than 80% in the medium and long run. The part of the error variance of the final goods output explained by supply shocks is 43% in the short run and more than 60% in the medium and long run. The part of GDP error variance determined by supply shocks varies from 50% to 70% at various horizons. This result confirms the conclusions of a baseline RBC model in which a business cycle is driven primarily by a technological shock. The result is also in line with structural VAR models with long-run restrictions in which the output is determined by a supply shock in the long run (Blanchard and Quah (1989)).

However, contrary to the identified VAR literature, the monetary policy shock is a source of macroeconomic volatility at all horizons. In the short run, the monetary policy shock accounts for 18.5% of the error variance of consumption, 38% of the error variance of GDP, 18.7% of the real exchange rate error variance, and 38.5% of the error variance of CPI inflation. In the long run, the importance of the monetary policy shock as a driving factor of an economy decreases, yet still remains significant. For example, monetary policy accounts for 17% of the GDP error variance at the 20-year horizon. The result is not a surprise: monetary policy explains even a larger part of long-term output error variance in the euro area (Smets and Wouters (2003)).

The preference shock is a primary driving force of consumption volatility in the short run, as it explains 45% of the error variance at the one-year horizon. In the medium and long run, consumption is driven mostly by supply shocks, like all measures of output.

The commodity export shock contributes much to GDP and import volatility at all horizons. It accounts for 23.8% of the error variance of GDP in the short run and about 19% in the long run. The portion of import volatility explained by the commodity export shock varies from 10.5% to 25.7%. It is noteworthy that the commodity export shock is not an important source of volatility of non-commodity output. The real appreciation induced by a positive commodity export shock increases imports and decreases exports. It is the reason why the consumption growth following a positive commodity export shock does not affect the intermediate goods output (the portion of error variance explained by the commodity export shock is close to zero at all horizons). Thus, the model shows symptoms of the Dutch disease in Russia at least before 2012.

The risk premium shock is the most important source of volatility of the real exchange rate in the short-run, accounting for 31.5% of the error variance, and, along with supply shocks, contributes significantly to the RER variance in the long run.

Contrary to existing literature, we do find that the mark-up shock explains a large part of the error variance of inflation. Vice versa, the impact of the price mark-up shock on all variables in the model, including prices, is not significant. It would be interesting to verify if this result is robust in the the case of another monetary policy rule or model setup. We leave this question for our future research.

Table 2. Forward error variance decomposition

Shock	C	Y	GDP	Q	Qd	Qex	Qim	$\mathcal{E}$	Pi	W
					1 year					
Preference	45.3	3.9	7.3	10.2	5.8	2.2	17.3	1.1	2.1	0.7
Labor supply	11	26.6	19.3	16.7	23.8	24.1	0.4	11.2	19.8	71.5
Commodity export	1.9	0.1	23.8	2.2	0.5	4.9	10.5	4.2	0.4	1.0
Technology	17.7	42.5	31.1	27.4	38.3	36.5	0.4	16.4	32.3	20.8
Monetary policy	18.5	26.5	38.0	34.0	29.8	2.1	18.6	18.7	38.5	0.1
Price mark-up	0.2	0.3	0.4	0.3	0.2	0.2	0.4	0.2	1.4	0.8
Risk premium	2.7	0.1	1.1	5.0	0.7	17.7	31.5	37.0	3.9	3.3
Foreign output	0	0	0.1	0	0	2.1	0.2	0.2	0	0
Foreign interest rate	2.7	0.1	1.6	4.1	0.8	10.1	20.8	11.0	1.6	1.9
Foreign inflation	0	0	0	0	0	0.1	0	0.1	0	0
					3 years					
Preference	22.0	2.0	4.0	5.5	3.0	1.2	11.3	0.8	2.0	0.5
Labor supply	28.6	42.3	35.0	31.7	39.7	34.2	0.6	20.9	22.3	71.9
Commodity export	4.8	0.3	19.0	4.4	1.1	7.5	22.7	6.3	0.6	1.7
Technology	27.3	41.8	34.4	31.2	39.2	34.0	0.5	20.5	31.9	21.0
Monetary policy	10.5	13.0	20.0	17.9	14.9	1.4	11.7	13.4	36.5	0.1
Price mark-up	0.1	0.1	0.2	0.2	0.2	0.1	0.3	0.1	1.3	0.4
Risk premium	1.9	0.1	1.1	3.1	0.5	9.9	22.7	26.8	3.8	1.9
Foreign output	0	0	0.1	0	0	1.7	0.5	0.3	0	0
Foreign interest rate	5.0	0.4	2.9	5.9	1.5	10.0	29.8	10.9	1.7	2.3
Foreign inflation	0	0	0	0	0	0.1	0	0.1	0	0
					20 years					
Preference	17.1	2.0	3.8	5.1	2.9	0.9	8.8	0.7	2.1	0.9
Labor supply	31.9	45.4	37.3	33.2	42.5	35.1	0.5	23.4	23.3	68.2
Commodity export	8.5	1.2	18.5	7.3	2.5	7.7	25.7	6.6	0.9	3.4
Technology	25.7	39.4	32.3	28.7	36.9	30.6	0.5	20.3	31.6	21.6
Monetary policy	7.9	10.9	16.6	14.7	12.4	1.4	8.4	11.5	35.3	0.2
Price mark-up	0.1	0.1	0.2	0.1	0.1	0.1	0.2	0.1	1.2	0.4
Risk premium	2.5	0.2	1.5	3.7	0.7	9.9	22.0	24.2	3.7	2.1
Foreign output	0.1	0.1	0.3	0.1	0	1.6	0.8	0.3	0	0.1
Foreign interest rate	6.3	0.8	3.8	7.2	2.0	12.7	33.2	12.7	1.8	3.1
Foreign inflation	0	0	0	0	0	0.1	0	0	0	0

Therefore, although Russia is an open economy, our results show that the fluctuations of macroeconomic variables are determined primarily by domestic shocks. Domestically based shocks account for 88% and 81% of the error variance of final goods output in the short run and long run, respectively. The only measure of economic activity that shows a considerable dependence on commodity dynamics is GDP because it explicitly accounts for export revenues. This result has some implications for macroeconomic policy. In the paper, we do not discuss an optimal monetary policy issue, but it does seem reasonable for policy makers to switch to inflation targeting in the near future, as the Central Bank of Russia promised to do by 2015.

1.4 Conclusion

In this paper, we constructed a DSGE model for an economy with commodity exports. The parameters of the model were estimated using Bayesian techniques on Russian data. Our principal goal was to identify the contribution of structural shocks to the business cycle fluctuations in an economy with commodity exports. Our main interest was the quantitative estimate of the impact of the commodity export shock on macroeconomic volatility in Russia. However, the model is general and may be estimated or calibrated for any export-oriented economy.

The paper is also an important step toward a general equilibrium model suitable for policy analysis and for forecasting similar models that are currently in use by central banks in many countries.

Our model yields plausible estimates, and the impulse response functions are in line with empirical evidence. We made a historical decomposition of two observed time series (consumption and real exchange rate) and one simulated time series to identify which shocks were the most influential in any particular quarter. It is interesting to note that the financial crisis of 2009 in Russia is captured by the model as a joint influence of risk premium shock and commodity export shock, which seems reasonable.

Finally, we determine the contribution of structural shocks to forecast error variance of endogenous variables in the short, medium, and long run.

Our results show that non-commodity output both for final and intermediate goods is determined by domestic demand (monetary policy) and supply shocks (shock of technology and labor supply shock) at all horizons. The commodity export shock does not contribute much to non-commodity output volatility, accounting for only 7.3% of the error variance of final goods output at the 20-year horizon. The likely reason is that the positive commodity shock results in real exchange rate appreciation, thereby decreasing exports and increasing imports. The commodity revenues shock accounts for up to 7.73% of the error variance of non-commodity exports and up to 25.71% of the error variance of imports in the long run. So the model shows the symptoms of the Dutch disease in Russia at least before 2012. However, commodity export revenues shock does contribute much to GDP, since GDP explicitly accounts for all export revenues. The shock accounts for 24% of the error variance of GDP in the short run and about 19% in the medium and long run. Consumption is driven primarily by preference shock in short run and by supply shocks in medium and long run. The most influential shocks for the real exchange rate are risk premium shock (at all horizons), monetary policy shock (in the short run) and supply shocks (in the medium and long run).

Our main conclusion is the following: in spite of a strong impact by commodity export shock on GDP, the business cycle in Russia is mostly domestically based. Although we do not explicitly consider an optimal monetary policy issue in the paper, the conclusion implies that it is reasonable for policy makers to switch to inflation targeting as the Central Bank of Russia is supposed to do by 2015.

We admit that our model may underestimate the impact of commodity exports on a domestic economy for two reasons. First, we do not split public and private consumption, so we do not account for an increase in government spending when the situation in the oil market is favorable. This could be crucial in the case of a higher propensity to spend in the public sector than in the private one. Second, the model is stationary and cannot account for permanent shocks. In this paper, we leave aside these possible extensions for computational reasons. Elaborating these issues is left for future research.

Appendices

.1 Normalization

$$p_t^d = \frac{P_t^d}{P_t} \qquad p_t^{ex} = \frac{P_t^{ex}}{P_t^f} \qquad p_t^{im} = \frac{P_t^{im}}{P_t}$$

$$r_t^K = \frac{R_t^K}{P_t} \qquad w_t = \frac{W_t}{P_t^f} \qquad mc_t = \frac{MC_t}{P_t}$$

$$\pi_t = \frac{P_t}{P_{t-1}} \qquad \pi_t^d = \frac{P_t^d}{P_{t-1}^d} \qquad \pi_t^f = \frac{P_t^f}{P_{t-1}^f}$$

$$\widetilde{p}_t^d = \frac{\widetilde{P}_t^d}{P_t} \qquad \widetilde{p}_t^{ex} = \frac{\widetilde{P}_t^{ex}}{P_t^f} \qquad \widetilde{p}_t^{im} = \frac{\widetilde{P}_t^{im}}{P_t}$$

$$\widetilde{w}_t = \frac{\widetilde{W}_t}{P_t} \qquad \mathcal{E}_t = \frac{S_t P_t^f}{P_t} \qquad d_t^f = \frac{D_t^f}{P_t^f}$$

$$o_t = \frac{O_t}{P_t^f}$$

.2 Steady-state derivation

$$p^d = (1+\upsilon)mc$$
$$p^{ex} = (1+\upsilon)\frac{mc}{\mathcal{E}}$$
$$p^{im} = (1+\upsilon)\mathcal{E}$$
$$p^d = (1+\upsilon)mc$$
$$p^{ex} = (1+\upsilon)\frac{mc}{\mathcal{E}}$$
$$1+i = \frac{1}{\beta}$$
$$1+i^* = \frac{1}{\beta}$$
$$\big(C(1-\nu)\big)^{-\sigma_1} = \mu$$
$$r^K = \frac{1}{\beta} - (1-\delta)$$
$$w = (1+\gamma)\big(C^{\sigma_1}(1-\nu)^{\sigma_1}\big)L^{\sigma_2}$$
$$A = 1$$
$$\begin{aligned} mc &= w^{1-\psi}\big(r^K\big)^{\psi}\psi^{-\psi}(1-\psi)^{\psi-1} = \\ &= ((1+\gamma)\left(C^{\sigma_1}(1-\nu)^{\sigma_1}\right)L^{\sigma_2})^{1-\psi}\,(r^K)^{\psi}\psi^{-\psi}(1-\psi)^{\psi-1} = \\ &= (1+\gamma)^{1-\psi}\left(\big(C^{\sigma_1}(1-\nu)^{\sigma_1}\big)L^{\sigma_2}\right)^{1-\psi}\left(r^K\right)^{\psi}\psi^{-\psi}(1-\psi)^{\psi-1} \end{aligned}$$
$$1 = \left(p^d\right)\left(p^{im}\right)^{\alpha_{im}} = ((1+\upsilon)mc)^{\alpha_d}((1+\upsilon)\mathcal{E})^{\alpha_{im}}$$
$$\mathcal{E} = (1+\upsilon)^{-\frac{1}{\alpha_{im}}}mc^{-\frac{\alpha_d}{\alpha_{im}}}$$

$$p^{ex} = (1+\upsilon)\frac{mc}{\mathcal{E}} = \frac{(1+\upsilon)mc}{(1+\upsilon)^{-\frac{1}{\alpha_{im}}}mc^{-\frac{\alpha_d}{\alpha_{im}}}} = (1+\upsilon)^{\frac{\alpha_{im}+1}{\alpha_{im}}}mc^{\frac{1}{\alpha_{im}}}$$
$$p^{im} = (1+\upsilon)\mathcal{E} = (1+\upsilon)^{1-\frac{1}{\alpha_{im}}}mc^{-\frac{\alpha_d}{\alpha_{im}}} = (1+\upsilon)^{-\frac{\alpha_d}{\alpha_{im}}}mc^{-\frac{\alpha_d}{\alpha_{im}}}$$

The current account equilibrium imply that:

$$p^{ex}Q^{ex} + o = \frac{p^{im}}{\mathcal{E}}Q^{im}$$

$$p^{ex}Q^{ex} = \alpha_{ex}Y^f$$

$$\alpha_{ex}Y_f + o = (1+\upsilon)Q^{im}$$

$$Q^{im} = \frac{\alpha_{ex}Y^f + o}{1+\upsilon}$$

$$Q^{im} = \alpha_{im}\frac{Q}{p^{im}}$$

$$\begin{aligned} Q = \frac{Q^{im}p^{im}}{\alpha_{im}} &= \frac{\alpha_{ex}Y^f + o}{1+\upsilon} \cdot \frac{(1+\upsilon)\mathcal{E}}{\alpha_{im}} = \\ &= \frac{1}{\alpha_{im}}(\alpha_{ex}Y^f + o)(1+\upsilon)^{-\frac{1}{\alpha_{im}}} mc^{-\frac{\alpha_d}{\alpha_{im}}} \end{aligned}$$

$$\frac{Q^d}{Q^{im}} = \frac{\alpha_d p_{im}}{\alpha_{im} p_d} = \frac{\frac{\alpha_d}{\alpha_{im}}(1+\upsilon)^{-\frac{\alpha_d}{\alpha_{im}}} mc^{-\frac{\alpha_d}{\alpha_{im}}}}{(1+\upsilon)mc} = \frac{\alpha_d}{\alpha_{im}}(1+\upsilon)^{-\frac{1}{\alpha_{im}}} mc^{-\frac{1}{\alpha_{im}}}$$

$$\begin{aligned} Q^d &= \frac{\alpha_{ex}Y^f + o}{1+\upsilon} \cdot \frac{\alpha_d}{\alpha_{im}}(1+\upsilon)^{-\frac{1}{\alpha_{im}}} mc^{-\frac{1}{\alpha_{im}}} = \\ &= (\alpha_{ex}Y^f + o)\frac{\alpha_d}{\alpha_{im}}(1+\upsilon)^{-1-\frac{1}{\alpha_{im}}} mc^{-\frac{1}{\alpha_{im}}} \end{aligned}$$

$$Q^{ex} = \alpha_{ex}Y^f(p^{ex})^{-1} = \alpha_{ex}Y^f\left((1+\upsilon)^{\frac{\alpha_{im}+1}{\alpha_{im}}} mc^{\frac{1}{\alpha_{im}}}\right)^{-1} = \alpha_{ex}Y^f(1+\upsilon)^{-\frac{\alpha_{im}+1}{\alpha_{im}}} mc^{-\frac{1}{\alpha_{im}}}$$

$$\begin{aligned} Y = Q^d + Q^{ex} &= (\alpha_{ex}Y^f + o)\frac{\alpha_d}{\alpha_{im}}(1+\upsilon)^{-\frac{\alpha_{im}+1}{\alpha_{im}}} mc^{-\frac{1}{\alpha_{im}}} + \alpha_{ex}Y^f(1+\upsilon)^{-\frac{\alpha_{im}+1}{\alpha_{im}}} mc^{-\frac{1}{\alpha_{im}}} = \\ &= \left(\alpha_{ex}Y^f + \frac{\alpha_d}{\alpha_{im}}\left(\alpha_{ex}Y^f + o\right)\right)(1+\upsilon)^{-\frac{\alpha_{im}+1}{\alpha_{im}}} mc^{-\frac{1}{\alpha_{im}}} = \alpha(1+\upsilon)^{-\frac{\alpha_{im}+1}{\alpha_{im}}} mc^{-\frac{1}{\alpha_{im}}} \end{aligned}$$

where $\alpha = \alpha_{ex}Y^f + \frac{\alpha_d}{\alpha_{im}}(\alpha_{ex}Y^f + o)$

Let us turn to labor and capital

$$\begin{aligned} K &= Y\left(\frac{\psi}{1-\psi} \cdot \frac{w}{r^K}\right)^{1-\psi} = \\ &= \alpha(1+\upsilon)^{-\frac{\alpha_{im}+1}{\alpha_{im}}} mc^{-\frac{1}{\alpha_{im}}}\left(\frac{\psi}{1-\psi} \cdot \frac{(1+\gamma)\left(C^{\sigma_1}(1-\nu)^{\sigma_1}\right)L^{\sigma_2}}{r^K}\right)^{1-\psi} \end{aligned}$$

We know that

$$mc = (1+\gamma)^{1-\psi} \left(C^{\sigma_1}(1-\nu)^{\sigma_1} L^{\sigma_2}\right)^{1-\psi} (r^K)^{\psi} \psi^{-\psi}(1-\psi)^{\psi-1}$$

so

$$\begin{aligned}
&(1+\gamma)^{1-\psi} \left(\left(C^{\sigma_1}(1-\nu)^{\sigma_1}\right) L^{\sigma_2}\right)^{1-\psi} = mc\big(r^K\big)^{-\psi} \psi^{\psi}(1-\psi)^{1-\psi} \Rightarrow \\
K &= \alpha(1+\upsilon)^{-\frac{\alpha_{im}+1}{\alpha_{im}}} mc^{-\frac{1}{\alpha_{im}}} \psi^{1-\psi}(1-\psi)^{\psi-1}\big(r^K\big)^{\psi-1} \times \\
&\times (1+\gamma)^{1-\psi}\big(C^{\sigma_1}(1-\nu)^{\sigma_1} L^{\sigma_2}\big)^{1-\psi} = \\
&= \alpha(1+\upsilon)^{-\frac{\alpha_{im}+1}{\alpha}{}_{im}} mc^{-\frac{1}{\alpha_{im}}} \psi^{1-\psi}(1-\psi)^{\psi-1}\big(r^K\big)^{\psi-1} mc\big(r^K\big)^{-\psi} \psi^{\psi}(1-\psi)^{1-\psi} = \\
&= \alpha\psi(1+\upsilon)^{-\frac{\alpha_{im}+1}{\alpha_{im}}} mc^{-\frac{\alpha_d}{\alpha_{im}}} \big(r^K\big)^{-1} \\
L &= K\frac{1-\psi}{\psi} \cdot \frac{r^K}{w} = (1+\upsilon)^{-\frac{\alpha_{im}+1}{\alpha_{im}}} mc^{-\frac{\alpha_d}{\alpha_{im}}} \frac{\alpha(1-\psi)}{(1+\gamma)(C^{\sigma_1}(1-\nu)^{\sigma_1} L^{\sigma_2}}
\end{aligned}$$

We know that:

$$\begin{aligned}
&(1+\gamma)C^{\sigma_1}(1-\nu)^{\sigma_1} L^{\sigma_2} = mc^{\frac{1}{1-\psi}} r^{K^{-\frac{\psi}{1-\psi}}} \psi^{\frac{\psi}{1-\psi}} (1-\psi) \Rightarrow \\
L &= \alpha(1-\psi)(1+\upsilon)^{-\frac{\alpha_{im}+1}{\alpha_{im}}} mc^{-\frac{\alpha_d}{\alpha_{im}}} \frac{1}{mc^{\frac{1}{1-\psi}} r^{K^{-\frac{\psi}{1-\psi}}} \psi^{\frac{\psi}{1-\psi}} (1-\psi)} = \\
&= \alpha(1+\upsilon)^{-\frac{\alpha_{im}+1}{\alpha_{im}}} mc^{-\frac{\alpha_d}{\alpha_{im}}-\frac{1}{1-\psi}} \big(r^K\big)^{\frac{\psi}{1-\psi}} \psi^{-\frac{\psi}{1-\psi}}
\end{aligned}$$

$$\begin{aligned}
&Y = K^{\psi} L^{1-\psi} = \\
&= \alpha \left(\psi(1+\upsilon)^{-\frac{\alpha_{im}+1}{\alpha_{im}}} mc^{-\frac{\alpha_d}{\alpha_{im}}} (r^K)^{-1}\right)^{\psi} \left((1+\upsilon)^{-\frac{\alpha_{im}+1}{\alpha_{im}}} mc^{-\frac{\alpha_d}{\alpha_{im}}-\frac{1}{1-\psi}} (r^K)^{\frac{\psi}{1-\psi}} \psi^{-\frac{\psi}{1-\psi}}\right)^{1-\psi} = \\
&= \alpha(1+\upsilon)^{-\frac{\alpha_{im}+1}{\alpha_{im}}} mc^{-\frac{1}{\alpha_{im}}}
\end{aligned}$$

Now we can determine steady state of consumption:

$$\begin{aligned}
C &= (1-\nu)^{-1}(1+\gamma)^{-\frac{1}{\sigma_1}} mc^{\frac{1}{(1-\psi)\sigma_1}} \big(r^K\big)^{-\frac{\psi}{1-\psi}\sigma_1} \psi^{\frac{\psi}{(1-\psi)\sigma_1}} (1-\psi)^{\frac{1}{\sigma_1}} L^{-\frac{\sigma_2}{\sigma_1}} = \\
&= (1-\nu)^{-1}(1+\gamma)^{-\frac{1}{\sigma_1}} mc^{\frac{1}{(1-\psi)\sigma_1}} \big(r^K\big)^{-\frac{\psi}{(1-\psi)\sigma_1}} \psi^{\frac{\psi}{(1-\psi)\sigma_1}} (1-\psi)^{\frac{1}{\sigma_1}} \times \\
&\times \left(\alpha(1+\upsilon)^{-\frac{\alpha_{im}+1}{\alpha_{im}}} mc^{-\frac{\alpha_d}{\alpha_{im}}-\frac{1}{1-\psi}} r^{K\frac{\psi}{1-\psi}} \psi^{-\frac{\psi}{1-\psi}}\right)^{\frac{-\sigma_2}{\sigma_1}} = \\
&= (1-\nu)^{-1}(1+\gamma)^{-\frac{1}{\sigma_1}} mc^{\frac{1}{(1-\psi)\sigma_1}} \big(r^K\big)^{-\frac{\psi}{(1-\psi)\sigma_1}} \psi^{\frac{\psi}{(1-\psi)\sigma_1}} (1-\psi)^{\frac{1}{\sigma_1}} \times \\
&\times \alpha^{-\frac{\sigma_2}{\sigma_1}} (1+\upsilon)^{(\frac{\alpha_{im}+1}{\alpha_{im}})\cdot\frac{\sigma_2}{\sigma_1}} mc^{\left(\frac{\alpha_d}{\alpha_{im}}+\frac{1}{1-\psi}\right)\cdot\frac{-\sigma_2}{\sigma_1}} r^{K(\frac{\psi}{1-\psi})\cdot\frac{(-\sigma_2)}{\sigma_1}} \psi^{\frac{\psi}{1-\psi}\cdot\frac{\sigma_2}{\sigma_1}} = \\
&= (1-\nu)^{-1}(1+\gamma)^{-\frac{1}{\sigma_1}} (1-\psi)^{\frac{1}{\sigma_1}} \psi^{\frac{\psi(1+\sigma_2)}{(1-\psi)\sigma_1}} \big(r^K\big)^{-\frac{\psi(1+\sigma_2)}{(1-\psi)\sigma_1}} \times \\
&\times \alpha^{-\frac{\sigma_2}{\sigma_1}} (1+\upsilon)^{\frac{\alpha_{im}+1}{\alpha_{im}}\cdot\frac{\sigma_2}{\sigma_1}} mc^{\frac{\alpha_d\sigma_2}{\alpha_{im}\sigma_1}+\frac{1+\sigma_2}{(1-\psi)\sigma_1}}
\end{aligned}$$

Steady-state of investment:

$$I = \delta K = \alpha\delta\psi(1+\upsilon)^{-\frac{\alpha_{im}+1}{\alpha_{im}}} mc^{-\frac{\alpha_d}{\alpha_{im}}} (r^K)^{-1}$$

Therefore, equation $Q = C + I$ takes the following form

$$\Upsilon_Q mc^{-\frac{\alpha_d}{\alpha_{im}}} = \Upsilon_C mc^{\frac{\alpha_d \sigma_2}{\alpha_{im}\sigma_1} + \frac{1+\sigma_2}{(1-\psi)\sigma_1}} + \Upsilon_I mc^{-\frac{\alpha_d}{\alpha_{im}}}$$

where

$$\begin{aligned}
\Upsilon_Q &= \frac{1}{\alpha_{im}}(\alpha_{ex}Y^f + o)(1+\upsilon)^{-\frac{1}{\alpha_{im}}} \\
\Upsilon_C &= (1-\nu)^{-1}(1+\gamma)^{-\frac{1}{\sigma_1}}(1-\psi)^{\frac{1}{\sigma_1}}\psi^{\frac{\psi(1+\sigma_2)}{(1-\psi)\sigma_1}} \times \\
&\times (r^K)^{-\frac{\psi(1+\sigma_2)}{(1-\psi)\sigma_1}} \alpha^{-\frac{\sigma_2}{\sigma_1}} (1+\upsilon)^{\frac{\alpha_{im}+1}{\alpha_{im}} \cdot \frac{\sigma_2}{\sigma_1}} \\
\Upsilon_I &= \alpha\delta\psi(1+\upsilon)^{-\frac{\alpha_{im}+1}{\alpha_{im}}}(r^K)^{-1}
\end{aligned}$$

We solve the equation $Q = C + I$:

$$\begin{aligned}
\Upsilon_C mc^{\frac{\alpha_d\sigma_2}{\alpha_{im}\sigma_1} + \frac{1+\sigma_2}{(1-\psi)\sigma_1} + \frac{\alpha_d}{\alpha_{im}}} &= \Upsilon_Q - \Upsilon_I \\
mc^{\frac{\alpha_d\sigma_2}{\alpha_{im}\sigma_1} + \frac{1+\sigma_2}{(1-\psi)\sigma_1} + \frac{\alpha_d}{\alpha_{im}}} &= \frac{\Upsilon_Q - \Upsilon_I}{\Upsilon_C} \\
mc &= \left(\frac{\Upsilon_Q - \Upsilon_I}{\Upsilon_C}\right)^{\left(\frac{\alpha_d\sigma_2}{\alpha_{im}\sigma_1} + \frac{1+\sigma_2}{(1-\psi)\sigma_1} + \frac{\alpha_d}{\alpha_{im}}\right)^{-1}} \\
&= \left(\frac{\Upsilon_Q - \Upsilon_I}{\Upsilon_C}\right)^{\frac{(1-\psi)\sigma_1\alpha_{im}}{\alpha_{im}+\sigma_2+\alpha_d\sigma_1-\alpha_d\psi\sigma_1-\alpha_d\psi\sigma_2}}
\end{aligned}$$

.3 Log-linearized model

$$\widehat{Q}_t^d = \widehat{Q}_t - \widehat{p}_d \tag{71}$$

$$\widehat{Q}_t^{im} = \widehat{Q}_t - \widehat{p}_{im} \tag{72}$$

$$\widehat{Q}_t^{ex} = \widehat{Y}_t^f - \widehat{p}_{ex} \tag{73}$$

$$\alpha_d \widehat{p}_t^d + \alpha_{im} \widehat{p}_t^{im} = 0 \tag{74}$$

$$\widehat{L}_t = \widehat{r}_t^K - \widehat{w}_t + \widehat{K}_t \tag{75}$$

$$\widehat{K}_t = -\widehat{A}_t + (1-\psi)\left(\widehat{w}_t - \widehat{r}_t^K\right) + \widehat{Y}_t \tag{76}$$

$$\widehat{m}c = -\widehat{A}_t + (1-\psi)\widehat{w}_t + \psi \widehat{r}_t^K \tag{77}$$

$$\begin{aligned} \widehat{p}_t^d - \theta_d \widehat{p}_{t-1}^d + \theta_d \widehat{\pi}_t = (1-\theta_d)(1-\beta\theta_d)\widehat{m}c_t + \\ + \beta\theta_d E_t(\widehat{p}_{t+1}^d - \theta_d \widehat{p}_t^d + \widehat{\pi}_{t+1}) \end{aligned} \tag{78}$$

$$\begin{aligned} \widehat{p}_t^{ex} - \theta_{ex} \widehat{p}_{t-1}^{ex} + \theta_{ex} \widehat{\pi}_t^f = (1-\theta_{ex})(1-\beta\theta_{ex})(\widehat{m}c_t - \widehat{\mathcal{E}}_t) + \\ + \beta\theta_{ex} E_t(\widehat{p}_{t+1}^{ex} - \theta_{ex} \widehat{p}_t^{ex} + \widehat{\pi}_{t+1}^f) \end{aligned} \tag{79}$$

$$\begin{aligned} \widehat{p}_t^{im} - \theta_{im} \widehat{p}_{t-1}^{im} + \theta_{im} \widehat{\pi}_t = (1-\theta_{im})(1-\beta\theta_{im})\widehat{\mathcal{E}}_t + \\ + \beta\theta_{ex} E_t(\widehat{p}_{t+1}^{im} - \theta_{im} \widehat{p}_t^{im} + \widehat{\pi}_{t+1}) \end{aligned} \tag{80}$$

$$\begin{aligned} \widehat{w}_t - \theta_w(\widehat{w}_{t-1} - \widehat{\pi}_t) = \frac{(1-\theta_w)(1-\beta\theta_w)\gamma}{(1+\gamma)\sigma_2 + \gamma}(\widehat{\epsilon}_t^l + \\ + \frac{1+\gamma}{\gamma}\sigma_2 \widehat{w}_t + \sigma_2 \widehat{L}_t + \frac{\sigma_1}{1-\nu}(\widehat{C}_t - \nu \widehat{C}_{t-1})) + \\ + \beta\theta_w(\widehat{w}_{t+1} - \theta_w(\widehat{w}_t - \widehat{\pi}_{t+1})) + \beta\theta_w E_t(1-\theta_w)\widehat{\pi}_{t+1} \end{aligned} \tag{81}$$

$$U_{c,t} = \widehat{\epsilon}_t^b - \frac{\sigma_1}{1-\nu}(\widehat{C}_t - \nu \widehat{C}_{t-1}) \tag{82}$$

$$E_t(\widehat{U}_{c,t+1} - \widehat{\pi}_{t+1}) + \widehat{\imath} = \widehat{U}_{c,t} \tag{83}$$

$$E_t(\widehat{U}_{c,t+1} + \widehat{\mathcal{E}}_{t+1} - \pi_{t+1}^f) + \widehat{\imath}^* = \widehat{U}_{c,t} + \widehat{\mathcal{E}}_t \tag{84}$$

$$\begin{aligned} \widehat{U}_{c,t} + \phi(\widehat{K}_{t+1} - \widehat{K}_t) = E_t(\widehat{U}_{c,t+1} + \beta r^K \widehat{r}_{t+1}^K + \\ + \beta\phi \widehat{K}_{t+2} - \beta\phi \widehat{K}_{t+1}) \end{aligned} \tag{85}$$

$$\widehat{K}_{t+1} = (1-\delta)\widehat{K}_t + \delta \widehat{I}_t \tag{86}$$

$$\widehat{\rho}_t = -\frac{\omega P^f}{P^{ex}Qex}\widehat{d}_t^f + \widehat{\epsilon}_\rho \tag{87}$$

$$\widehat{\imath} = z_1\widehat{\imath}_{t-1} + (1-z_1)z_2\widehat{\pi}_t + (1-z_1)z_3\widehat{Y}_t + \epsilon_z; \tag{88}$$

$$\widehat{Y}_t = \frac{Q^d}{Y}\widehat{Q}_t^d + \frac{Q^{ex}}{Y}\widehat{Q}_t^{ex} \tag{89}$$

$$\widehat{Q}_t = \frac{C}{Q}\widehat{C}_t + \frac{I}{Q}\widehat{I}_t \tag{90}$$

$$\begin{aligned} p^{ex}Q^{ex}(\widehat{p}_t^{ex} + \widehat{Q}_t^{ex}) + o\widehat{o}_t - \frac{p^{im}Q^{im}}{\mathcal{E}}(\widehat{p}_t^{im} + \widehat{Q}_t^{im} - \widehat{\mathcal{E}}_t) - \\ -\widehat{d}_t^f + \widehat{\imath}^*\widehat{d}_{t-1}^f = 0 \end{aligned} \tag{91}$$

$$\begin{aligned} \widehat{gdp}_t = \widehat{Q}_t + \frac{\mathcal{E}p^{ex}Q^{ex}}{Q}(\widehat{\mathcal{E}}_t + \widehat{p}_t^{ex} + \widehat{Q}_t^{ex}) + \frac{\mathcal{E}o}{Q}(\widehat{\mathcal{E}}_t + \widehat{o}_t) - \\ -\frac{p^{im}Q^{im}}{Q}(\widehat{p}_t^{im} + \widehat{Q}_t^{im}) \end{aligned} \tag{92}$$

$$\widehat{\imath}^* = \widehat{\imath}_t^f + \widehat{\rho}_t \tag{93}$$

$$\widehat{\pi}_d = \widehat{p}_t^d - \widehat{p}_{t-1}^d + \widehat{\pi} \tag{94}$$

$$\widehat{\epsilon}_t^b = \rho_b\widehat{\epsilon}_{t-1}^b + \varepsilon_t^b \tag{95}$$

$$\widehat{\epsilon}_t^l = \rho_l\widehat{\epsilon}_{t-1}^l + \varepsilon_t^l \tag{96}$$

$$\widehat{A}_t = \rho_a\widehat{A}_{t-1} + \varepsilon_t^A \tag{97}$$

$$\widehat{\imath}_t^f = \rho_{if}\widehat{\imath}_{t-1}^f + \varepsilon_t^{if} \tag{98}$$

$$\widehat{\pi}_t^f = \rho_{\pi_f}\widehat{\pi}_{t-1}^f + \varepsilon_t^{\pi_f} \tag{99}$$

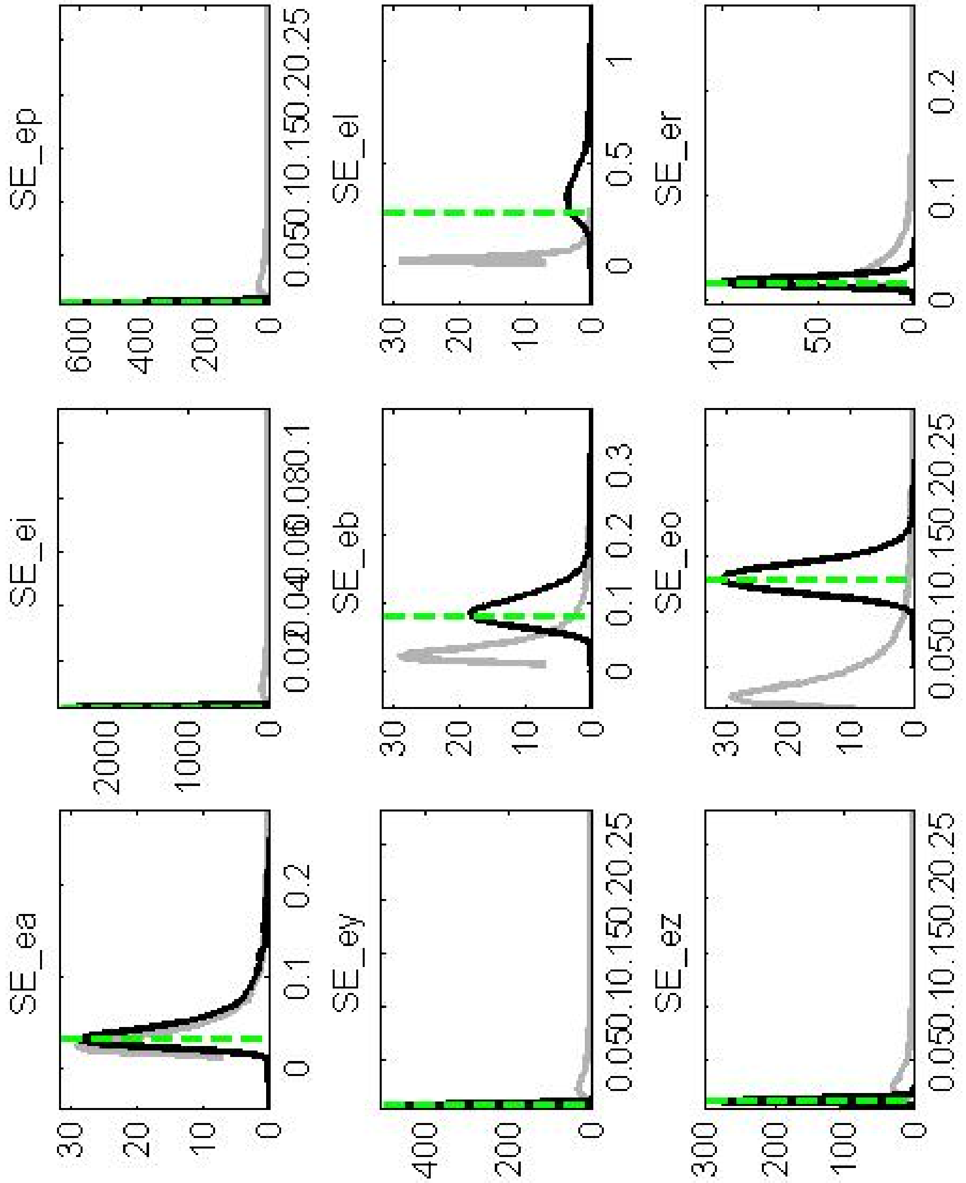
$$\widehat{Y}_t^f = \rho_{Yf}\widehat{Y}_{t-1}^f + \varepsilon_t^{Y_f} \tag{100}$$

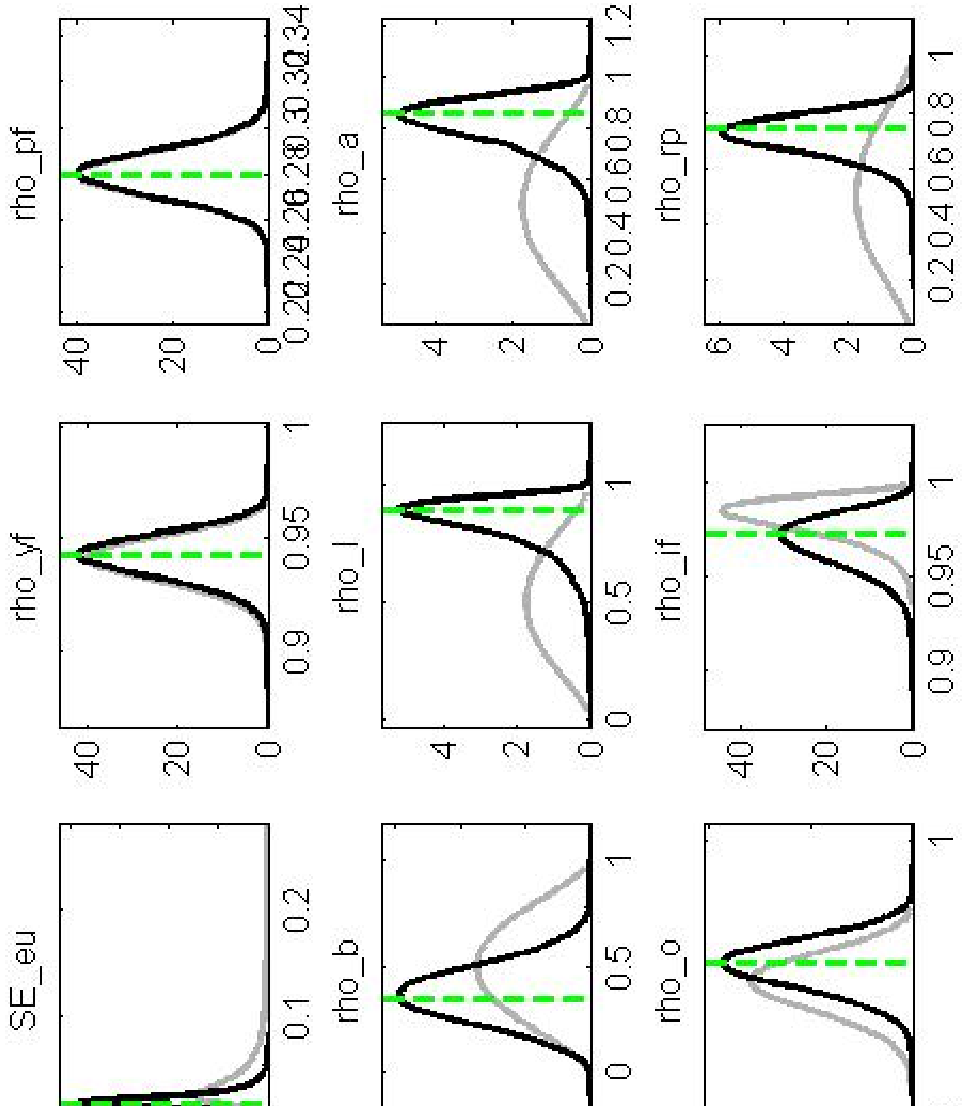
$$\widehat{o}_t = \rho_o\widehat{o}_{t-1} + \varepsilon_t^o \tag{101}$$

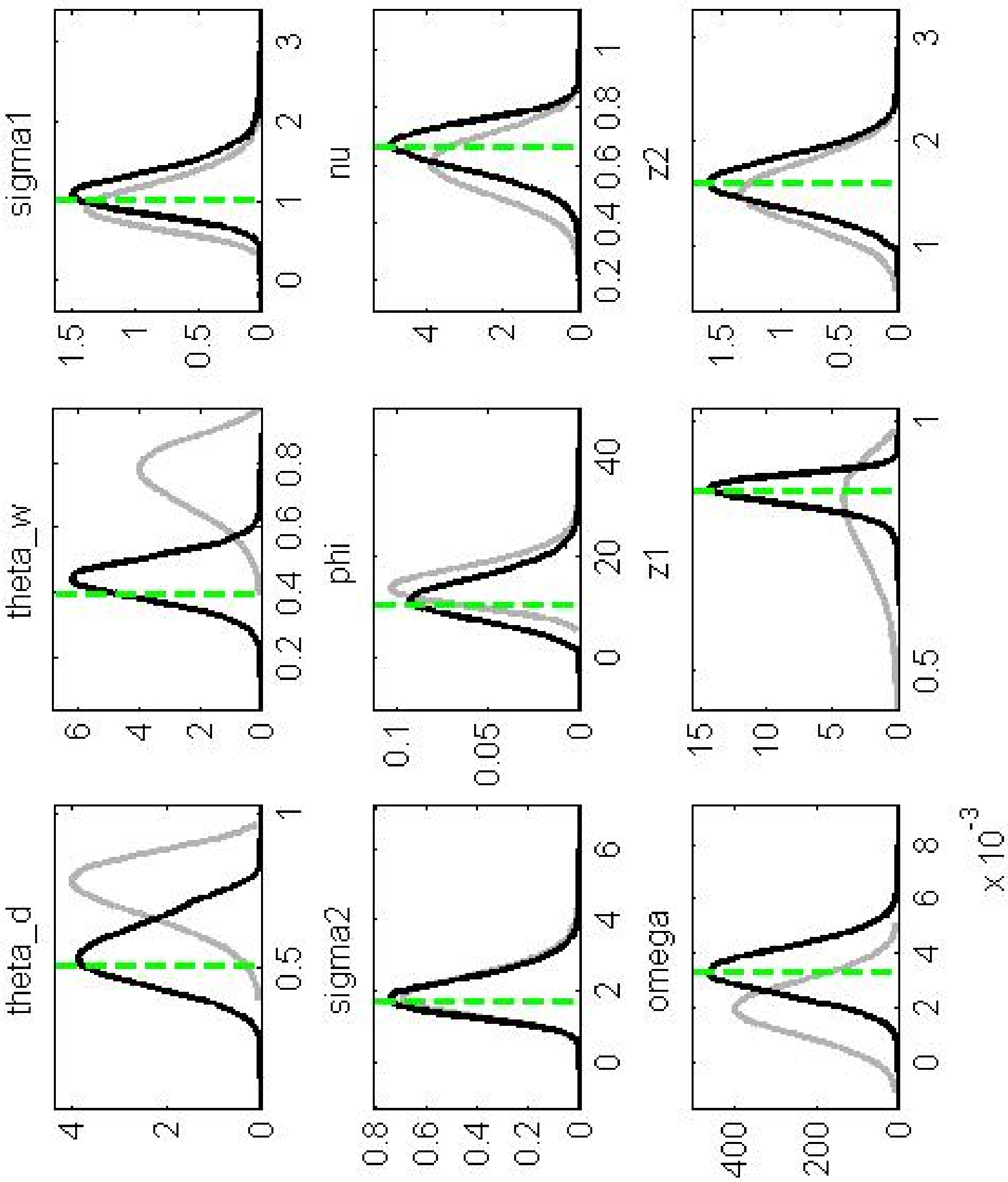
$$\widehat{\epsilon}_{\rho,t} = \rho_\rho\widehat{\epsilon}_{\rho,t-1} + \varepsilon_t^\rho \tag{102}$$

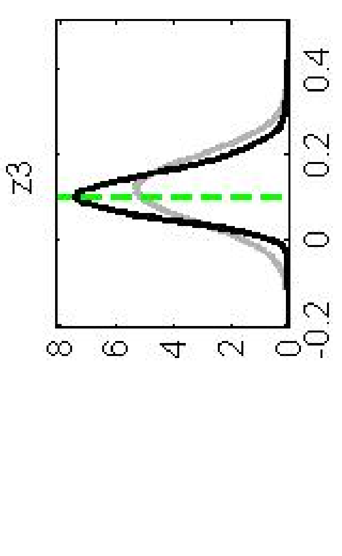
$$\tag{103}$$

.4 Priors and posteriors




SE_eu
rho_yf
rho_pf
rho_b
rho_l
rho_a
rho_o
rho_if
rho_rp


theta_d
theta_w
sigma1
sigma2
phi
nu
omega
z1
z2
x 10^-3



z3
8
6
4
2
0
-0.2
0
0.2
0.4

Chapter 2

BVAR Mapping[1]

2.1 Introduction

Precise forecasts are extremely important for macroeconomic policy. Due to policy lags, fiscal and monetary policy makers need to base their policy measures on forecasts and not on the actual values, as decisions taken today affect the economy only after a certain period of time. Therefore, an accurate forecast of macroeconomic indicators is a crucial factor of a successful policy.

Currently, a main workhorse for macroeconomic time series forecasting is a vector autoregression (VAR) model and its modifications. VAR models have become widely spread in macroeconomic analysis due to a critique aimed at previously commonly used traditional econometric models . For instance, Sims (1980) blamed incredible restrictions accepted ad hoc in the traditional framework,[2] and suggested that the researchers use a VAR model as a simply formulated dynamic model based on Wold decomposition that did not require any additional restrictions on the joint dynamics of included variables.

VAR models are widely used both for forecasting and structural analysis due to their self-consistency and relative simplicity. However, to reflect correctly the dynamics of actual time series, a VAR model requires many lags that may lead to over-parametrization,

[1]co-authored with Boris Demeshev, NRU HSE; published as Demeshev and Malakhovskaya (2016)

[2]We call 'traditional models' the models in vein of the the Cowles comission approach. Their forecasting performance dropped abrubtly in the beginning of 1970s, that is approximately at the same time as the traditional Phillips curve "disappeared" (for futher detail see Favero (2001) and Malakhovskaya and Pekarsky (2012))

inefficient estimation, and high forecast errors. Another important question concerning a VAR model is how many variables to include. Using many variables is motivated by the fact that currently central banks of developed countries monitor a large number of indicators, and a small-dimensional VAR model cannot reflect all the information that is available for central banks. Hence, using high-dimensional models may potentially increase the forecasting accuracy. However, the increase of the number of variables in the model exacerbates the problem of over-parametrization, non-efficient estimation and high forecast errors. One of the solutions to this problem is shrinking the parameters to certain values by imposing some prior information in the form of prior distributions on parameters and errors covariance matrix, or in other words, using Bayesian VAR (BVAR) instead of traditional (or frequentist) VAR. Another widely-known VAR-based solutions to over-parametrization problem are DFM and FAVAR models (see Geweke, 1977 and Bernanke, Boivin, and Eliasz, 2005 for pioneer papers and Stock and Watson, 2016 for a review) that are left out of the scope of this chapter.

Researchers distinguish two key advantages of BVAR models comparing to the frequentist ones. First, this class of models proposes a solution to a over-parametrization problem and consequently permits including more variables into a model. At the same time, prior beleifs decrease incertainty in distributions of model parameters and increase the model's forecasting performance. Second, the prior distributions that are currently commonly used in applied research reflect the contemporaneous beleifs about long-run dynamics of the variables. This long-run dynamics cannot be detected in short samples that are usually available for research. Imposing prior distributions also increases the forecasting accuracy of the model. Besides, contemporanous computers realize simulations so rapidly that researchers are no more limited to make use of conjugate distributions only (the distributions for which prior, likelyhood function and posterior appertain to the same class) that allow for a closed-form solution. It definitly increases the attractiveness of the Bayesian approach and fosters its expansion in macroeconomic analysis. (Karlsson (2013)).

The most often noticed disadvantage of the Bayesian approach is subjectivity as a model embodies the beliefs of its author. However, we think that this disadvantage is not really essential. As a matter of fact, a change of a prior distribution affects the results of the estimation[3]. At the same time, a frequentist VAR (and any other econometric model as well) is also a reflection of subjective beliefs of the researcher. A choice of a limited number of variables in the model, distinguishing exogenous and endogenous variables, a choice of a number of lags in the model [4] etc in any case reflect the researcher's personal beliefs about a proper model specification. Contrary to frequentist VAR models, the Bayesian VARs determine the subjectivity explicitly with prior distributions. Unfortunately, despite a widespread use of BVARs in research papers, the reviews of this approach are rare. The reviews by Karlsson (2013), Del Negro and Schorfheide (2011) and Canova (2007) are mathematically intensive and hardly comprehensible for economists without special mathematical education. Moreover, the reviews do not contain a very detailed classification of prior distributions and most of them do not contain a guide to realize the methods in an econometric software. The exceptions are Koop and Korobilis (2010) and Blake and Mumtaz (2012), that are accompanied by MATLAB codes.[5] However, Koop and Korobilis (2010) do not discuss a method of defining priors with dummy observations that has become very popular recently, including sum-of-coefficients prior and initial observation prior. Blake and Mumtaz (2012) use some non-conventional terminology, and their code contains an example of BVAR estimation with Gibbs sampler only (even where it is not needed). No review contains a detailed discussion of forecasting techniques with BVAR though the forecasting is a principal reason of the estimation of a BVAR model in a reduced form. This review contains a detailed classification of priors that are the most popular in macroeconomic applied research as well as forecasting techniques description. The paper is accompanied by a package in R that uses the same notations that the text that can be freely used both for study and research. We do not discuss here structural BVAR models (SBVAR), BVAR with time-varying parameters (TVP-BVAR), BVAR with stochastic volatility, and the choice of variable for a BVAR model estimation (see Uhlig, 1997, Koop and Korobilis, 2010, and Del Negro and Schorfheide, 2011 for reviews of these

[3]The estimation results change as a posterior density is a combination of ta prior density and a likelihood function. This question is thoroughly discussed in the next section.

[4]In practice to choose the number of lags in a frequentist VAR researchers use information criteria. However, different information criteria regularly suggest different numbers of lags. If it happens, a choice of a reliable criteria is a matter of a personal subjective decision.

[5]The codes in open access that we are aware of are described in Appendix 1.

methods).

2.2 BVAR estimation

2.2.1 Bayesian VAR: a model framework

Let y_{it}, be variables in a vector $y_t = (y_{1t}, y_{2t}, \ldots, y_{mt})'$ of dimension m [6] A VAR model in a reduced form is represented as:

$$y_t = \Phi_{const} + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \ldots + \Phi_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma) \tag{2.1}$$

with $\Phi_{const} = (c_1, \ldots, c_m)'$ being a constant vector of a dimension m, Φ_l being autoregression matrices of a dimension $m \times m$ with $l = 1, \ldots, p$. A vector ε_t is a vector of errors of a dimension m uncorrelated with regressors. Grouping the matrices into one combined matrix $\Phi = [\Phi_1 \ldots \Phi_p \; \Phi_{const}]'$ and defining a new vector $x_t = [y'_{t-1} \ldots y'_{t-p} \; 1]'$, gives a more compact way to represent a VAR model:

$$y_t = \Phi' x_t + \varepsilon_t \tag{2.2}$$

Grouping the variables and shocks in the following way : $Y = [y_1, y_2, \ldots, y_T]'$, $X = [x_1, x_2, \ldots, x_T]'$, $E = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T]'$, gives a matrix form of the VAR model:

$$Y = X\Phi + E \tag{2.3}$$

The same model can also be written in a vectorized form[7]:

$$\vec{(Y)} = \vec{(X\Phi I)} + \vec{(E)} \Leftrightarrow \tag{2.4}$$

$$y = (I_M \otimes X)\phi + \varepsilon \tag{2.5}$$

with $\varepsilon \sim \mathcal{N}(0, \Sigma \otimes I_T)$ and a vector $\phi = \vec{\Phi}$ being of a dimension $km \times 1$.

A key step of the Bayesian estimation is to find posterior distributions of the model parameters $p(\Phi, \Sigma|Y)$ using a likelihood function $p(Y|\Phi, \Sigma)$ and a given prior distribution, $p(\Phi, \Sigma|Y)$ according to the Bayes rule:

$$p(\Phi, \Sigma|Y) = \frac{p(\Phi, \Sigma) p(Y|\Phi, \Sigma)}{p(Y)} \tag{2.6}$$

[6]For readers' convenience all notations are also given in Appendix 2.

[7]Equation(2.5) of the system follows from the identity : $\vec{(ABC)} = (C \otimes A)\vec{(B)}$

As $p(Y)$ does not depend on Φ и Σ, the following expression is valid:

$$p(\Phi, \Sigma | Y) \propto p(\Phi, \Sigma) p(Y | \Phi, \Sigma) \tag{2.7}$$

As $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$, then the likelihood function is defined as :[8]

$$p(Y | \Phi, \Sigma) \propto |\Sigma|^{-T/2} \operatorname{etr} \left\{ -\frac{1}{2} \left[\Sigma^{-1} (Y - X\Phi)'(Y - X\Phi) \right] \right\} \tag{2.8}$$

Another way to write the likelihood function is:

$$p(Y | \Phi, \Sigma) \propto |\Sigma|^{-T/2} \operatorname{etr} \left\{ -\frac{1}{2} \left[\Sigma^{-1} \hat{E}' \hat{E} \right] \right\} \times \operatorname{etr} \left\{ -\frac{1}{2} \left[\Sigma^{-1} (\Phi - \hat{\Phi})' X' X (\Phi - \hat{\Phi}) \right] \right\}, \tag{2.9}$$

where $\hat{E} = Y - X\hat{\Phi}$ и $\hat{\Phi} = (X'X)^{-1} X'Y$.

In two next sections we discuss the most commonly used prior and posterior distributions.

2.2.2 Model estimation with different priors

Minnesota prior

A Bayesian solution to the over-identification problem was proposed by Litterman (1979), in the same paper Litterman shows that restrictions in the form of prior distributions increase the accuracy of estimates and forecasts. A prior distribution called the «Minnesota prior» was introduced by Litterman (1986) and (with some modifications) by Doan, Litterman, and Sims (1984).

The prior distribution is assumed to be multivariate normal and depending on several hyperparameters. The stochastic processes governing the parameters are assumed to be independent, and therefore the covariance matrix of the parameter vector ϕ noted as Ξ is diagonal. The covariance matrix of the vector ε_t that we note as Σ is also assumed to be diagonal and constant. Therefore, the vector ϕ does not depend on Σ:

$$\phi \sim \mathcal{N}(\underline{\phi}, \Xi) \tag{2.10}$$

The prior distribution density of ϕ can be written as:

$$p(\phi) = \frac{1}{(2\pi)^{km/2} |\Xi|^{1/2}} \exp \left\{ -\frac{1}{2} (\phi - \underline{\phi})' \Xi^{-1} (\phi - \underline{\phi}) \right\}. \tag{2.11}$$

[8]Here and below $\operatorname{etr}(\cdot) = \exp(\operatorname{tr}(\cdot))$.

Combining it with the likelihood function (2.8), we get that posterior distribution of parameters is given by the following equation:

$$\phi|Y \sim \mathcal{N}(\overline{\phi}, \overline{\Xi}) \tag{2.12}$$

with

$$\overline{\Xi} = [\underline{\Xi}^{-1} + \Sigma^{-1} \otimes (X'X)]^{-1}$$
$$\overline{\phi} = \overline{\Xi}[\underline{\Xi}^{-1}\underline{\phi} + (\Sigma^{-1} \otimes X')y].$$

If $\underline{\Xi}$ has a structure of the Kronecker product: $\underline{\Xi} = \Sigma \otimes \underline{\Omega}$, then the formula (2.12) can be simplified.

$$\overline{\Xi} = [\underline{\Xi}^{-1} + \Sigma^{-1} \otimes (X'X)]^{-1} = [(\Sigma \otimes \underline{\Omega})^{-1} + \Sigma^{-1} \otimes (X'X)]^{-1} =$$
$$= [\Sigma^{-1} \otimes \underline{\Omega}^{-1} + \Sigma^{-1} \otimes (X'X)]^{-1} = \Sigma \otimes (\underline{\Omega}^{-1} + X'X)^{-1} = \Sigma \otimes \overline{\Omega} \tag{2.13}$$

The derivation permits to decrease the dimension of the matrices to be inverted.

As a result we get:

$$\Phi|Y \sim \mathcal{N}(\overline{\Phi}, \Sigma \otimes \overline{\Omega}) \tag{2.14}$$

In practice instead of the matrix Σ its estimate $\hat{\Sigma}$ is used. The diagonal elements of $\hat{\Sigma}$ are equal to: $\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_m^2$, with $\hat{\sigma}_i^2$ being an error variance estimate in the $AR(p)$ model fitted for the series i. Some researchers use AR(1) model to compute the error variance estimate even the VAR model in hand has more lags.

The prior mean of the parameters can be written using a matrix $\underline{\Phi} = \mathbb{E}(\Phi)$ of dimension $k \times m$, with $\underline{\Phi} = [\underline{\Phi}_1 \ldots \underline{\Phi}_p \ \underline{\Phi}_{const}]'$ and $\underline{\phi} = \vec{\underline{\Phi}}$.

$$(\underline{\Phi}_l)_{ij} = \begin{cases} \delta_i & \text{if } i = j, l = 1; \\ 0, & \text{in other cases} \end{cases} \tag{2.15}$$

The Minnesota prior embodies the non-stationarity of macroeconomic time series. Currently, a common practice is to set $\delta_i = 1$ for non-stationary series and $\delta_i < 1$ for stationary series.

The Minnsota prior assumes that the prior covariance matrix of parameters $\underline{\Xi}$ is diagonal. The main diagonal of the matrix $\underline{\Xi}$ is split into m blocks $\underline{\Xi}_1, \underline{\Xi}_2, \ldots, \underline{\Xi}_m$ of dimension $k \times k$. In its turn, each block $\underline{\Xi}_i$, $i = 1, \ldots, m$ can be split into diagonal sub-blocks of dimension $m \times m$: $\underline{\Xi}_{i,lag=l}$, $l = 1, \ldots, p$ with a scalar $\underline{\Xi}_{i,const}$ in the end of the main diagonal:

$$\underline{\Xi} = \begin{pmatrix} \underline{\Xi}_1 & 0_{k\times k} & \cdots & 0_{k\times k} & 0_{k\times k} \\ 0_{k\times k} & \underline{\Xi}_2 & \cdots & 0_{k\times k} & 0_{k\times k} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{k\times k} & 0_{k\times k} & \cdots & \underline{\Xi}_{m-1} & 0_{k\times k} \\ 0_{k\times k} & 0_{k\times k} & \cdots & 0_{k\times k} & \underline{\Xi}_m \end{pmatrix} \quad \underline{\Xi}_i = \begin{pmatrix} \underline{\Xi}_{i,lag=1} & 0_{m\times m} & \cdots & 0_{m\times m} & 0_{m\times 1} \\ 0_{m\times m} & \underline{\Xi}_{i,lag=2} & \cdots & 0_{m\times m} & 0_{m\times 1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{m\times m} & 0_{m\times m} & \cdots & \underline{\Xi}_{i,lag=p} & 0_{m\times 1} \\ 0_{1\times m} & 0_{1\times m} & \cdots & 0_{1\times m} & \underline{\Xi}_{i,const} \end{pmatrix}$$

The diagonal elements of $\underline{\Xi}_i$ noted $\underline{\Xi}_{i,lag=l}$ are defined according to the formula:

$$(\underline{\Xi}_{i,lag=l})_{jj} = \begin{cases} \left(\frac{\lambda_{tight}}{l^{\lambda_{lag}}}\right)^2, & j = i \\ \left(\frac{\lambda_{tight}\cdot\lambda_{kron}\sigma_i}{l^{\lambda_{lag}}\sigma_j}\right)^2, & j \neq i \end{cases} \qquad \underline{\Xi}_{i,const} = \lambda_{tight}^2\lambda_{const}^2\sigma_i^2, \tag{2.16}$$

As equation (2.16) shows, the prior variance of the parameters depends on several hyperparameters set by the researcher. The hyperparameters express different features of the prior distribution. [9]

λ_{tight} (the shrinkage parameter) reflects the overall tightness of the prior distribution.

If $\lambda_{tight} \to 0$, than the prior distribution coincides with the posterior distribution and the data are not involved in the parameter estimation. In this case the prior is so tight as if the researcher pretends to know the parameters for sure, so that:

$$\Phi \sim \mathcal{N}(\underline{\Phi}, 0), \;\; \Phi|Y \sim \mathcal{N}(\underline{\Phi}, 0)$$

If $\lambda_{tight} \to \infty$, than the posterior mean of the parameters converges to the OLS estimate. In this case $\underline{\Xi}^{-1} = 0$, and so

$$\overline{\Xi} = (\Sigma^{-1} \otimes (X'X))^{-1} = \Sigma \otimes (X'X)^{-1}$$

[9] The relation between hyperparameters defined in different papers can be found in Appendix 3.

It leads to:

$$\overline{\phi} = 0 + (\Sigma \otimes (X'X)^{-1}) \cdot (\Sigma^{-1} \otimes X') \cdot y = (I \otimes (X'X)^{-1}X') \cdot \vec{(}Y) = \\ = \vec{(}(X'X)^{-1}X'Y \cdot I') = \vec{(}(X'X)^{-1}X'Y) \quad (2.17)$$

A cross-shrinkage parameter λ_{kron} gives some additional tightness to lags of all variables besides the dependent one in each equation. If $\lambda_{kron} < 1$, than the own lags of the dependent variable are assumed to forecast the value of the variable better than lags of other variables in the system. Therefore, the coefficients on other variables' lags are closer shrunk to zero.

If $\lambda_{kron} = 1$ the matrix $\underline{\Xi}$ has a Kronecker product structure and can be written as follows:

$$\underline{\Xi} = \Sigma \otimes \underline{\Omega},$$

with $\underline{\Omega}$ being a matrix of dimension $k \times k$, corresponding to a separate equation. Equation (2.2.2) means that the variances given in the matrix $\underline{\Omega}$ are multiplied by coefficient σ_i^2. The matrix $\underline{\Omega}$ can be written as:

$$\underline{\Omega} = \begin{pmatrix} \underline{\Omega}_{lag=1} & 0_{m\times m} & \cdots & 0_{m\times m} & 0_{m\times 1} \\ 0_{m\times m} & \underline{\Omega}_{lag=2} & \cdots & 0_{m\times m} & 0_{m\times 1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{m\times m} & 0_{m\times m} & \cdots & \underline{\Omega}_{lag=p} & 0_{m\times 1} \\ 0_{1\times m} & 0_{1\times m} & \cdots & 0_{1\times m} & \underline{\Omega}_{const} \end{pmatrix} \quad (2.18)$$

The matrix $\underline{\Omega}_{lag=l}$ has a dimension $m \times m$, and its diagonal elements are defined according to the formula:

$$(\underline{\Omega}_{lag=l})_{jj} = \left(\frac{\lambda_{tight}}{l^{\lambda_{lag}}\sigma_j}\right)^2 \quad \underline{\Omega}_{const} = \lambda_{tight}^2 \lambda_{const}^2 \quad (2.19)$$

The parameter λ_{const} reflects the a relative tightness of the prior of the constant, and parameter λ_{lag} reflects the speed of decrease of the prior variance when the number of lags increases.

As said above, if the Minnesota prior is applied, the Gibbs algorithm is not needed to compute the posterior. A random sample from the posterior can be obtained using Monte-Carlo simulations:

$$\phi^{[s]} \sim \mathcal{N}(\overline{\phi}; \Xi), \tag{2.20}$$

where $\phi^{[s]}$ is a realization of the parameter vector at $s-th$ step of the simulations.

If Ξ has a Kronecker product structure: $\underline{\Xi} = \Sigma \otimes \underline{\Omega}$, than instead of the vector $\phi^{[s]}$ it is possible to generate the matrix $\Phi^{[s]}$ numerically following a simpler algorithm:

1. Generate a matrix V of the dimension $k \times m$ using independent standard normal random values

2. Compute the matrix $\Phi^{[s]}$ according to the formula:

$$\Phi^{[s]} = \overline{\Phi} + \mathrm{chol}(\overline{\Omega}) \cdot V \cdot \mathrm{chol}(\Sigma^{[s]})', \tag{2.21}$$

with $\mathrm{chol}(\overline{\Omega})$ and $\mathrm{chol}(\Sigma^{[s]})$ being the upper triangular matrices resulted from the Cholesky decomposition of the matrices $\overline{\Omega}$ and $\Sigma^{[s]}$, respectively.

We can name several advantages of the Minnesota prior. First, it is easily defined. It has been successfully applied to tackle varied problems in economic research. As the posterior distribution is normal, then any parameter function can be computed easily using Monte-Carlo methods. An essential drawback of this prior is that it does not imply any Bayesian procedure to estimate the covariance matrix Σ.

Independent normal - inverse Wishart distribution

A generalization of the Minnesota prior is the independent normal - inverse Wishart prior (iNIW prior). It relaxes the constant covariance matrix of parameters assumption and can be written as:

$$\begin{cases} \phi \sim \mathcal{N}(\underline{\phi}; \Xi) \\ \Sigma \sim \mathcal{IW}(\underline{S}; \underline{\nu}) \\ \phi \text{ and } \Sigma \text{ are independent} \end{cases} \tag{2.22}$$

The Minnesota prior represents a particular case of the iNIW prior for $\underline{S} = (\underline{\nu} - m - 1) \cdot \Sigma$ and $\nu \to \infty$. In case of NIW prior it can be shown (Karlsson (2013)) that conditional posterior distributions can be written as:

$$\begin{cases} \phi|\Sigma, Y \sim \mathcal{N}(\overline{\phi}; \overline{\Xi}) \\ \Sigma|\phi, Y \sim \mathcal{IW}(\overline{S}; \overline{\nu}) \end{cases} \tag{2.23}$$

with the posterior hyperparameters defined as:

$$\begin{aligned}
&\overline{\nu} = \underline{\nu} + T \\
&\overline{S} = \underline{S} + E'E, \text{ где } E = Y - X\Phi \\
&\overline{\Xi} = (\underline{\Xi}^{-1} + \Sigma^{-1} \otimes X'X)^{-1} \\
&\overline{\phi} = \overline{\Xi} \cdot (\underline{\Xi}^{-1}\underline{\phi} + \vec{(}X'Y\Sigma^{-1})) = \\
&= \overline{\Xi} \cdot (\underline{\Xi}^{-1}\underline{\phi} + (\Sigma^{-1} \otimes (X'X))\vec{\hat{\Phi}})
\end{aligned}$$

The hyperparameters of the iNIW prior distribution can be chosen in the same way as for the Minnesota prior ((3.6) and (2.16)). If necessary a diffuse prior distribution for coefficients at predetermined variables can by defined by setting the respective value in the matrix $\underline{\Xi}^{-1}$ to zero.

If an arbitrary error covariance matrix is used, only conditional posterior distributions for ϕ и Σ are known. The Gibbs algorithm is required to get realizations from the joint posterior distribution.

The following steps produce a Markov chain that converges to the posterior distribution :

Step 1. Generate an arbitrary initial matrix $\Sigma^{[0]}$, for example, a unity matrix

Step 2. At step s generate ϕ and Σ according to:

$$\phi^{[s]} \sim \mathcal{N}(\overline{\phi}^{[s-1]}; \overline{\Xi}^{[s-1]}), \text{ where } \overline{\phi}^{[s-1]} \text{ and } \overline{\Xi}^{[s-1]} \text{ are computed with } \Sigma^{[s-1]} \tag{2.24}$$

$$\Sigma^{[s]} \sim \mathcal{IW}(\overline{S}^{[s]}; \overline{\nu}), \text{ where} \overline{S}^{[s]} \text{ are computed with } \phi^{[s]} \tag{2.25}$$

conjugate normal - inverse Wishart prior distribution

As it is said above, a drawback of the Minnesota prior is that the covariance matrix Σ is not estimated using some Bayesian procedures. To overcome this drawback, a conjugate prior can be used.

Given the covariance matrix of errors, the likelihood function can be split into two parts. One of them is proportional to normal distribution, and the another one is proportional to the inverse Wishart distribution. Therefore, the conjugate prior for the model is also the normal - inverse Wishart prior.

The conjugate normal - inverse Wishart prior distribution can be written as:

$$\begin{cases} \Sigma \sim \mathcal{IW}(\underline{S}, \underline{\nu}) \\ \phi|\Sigma \sim \mathcal{N}(\underline{\phi}, \Sigma \otimes \underline{\Omega}) \end{cases} \tag{2.26}$$

Contrary to the Minnesota prior, the conjugate normal - inverted Wishart prior (3.5) is always written for the case, when the parameter covariance matrix has a Kronecker structure, that is λ_{kron} is assumed to be equal to unity.

The hyperparameters of the mathematical expectation vector ($\underline{\phi}$) and covariance matrix $\underline{\Omega}$ of conditional prior distribution can be set as for the Minnesota prior for a particular case of $\lambda_{kron} = 1$ (see (3.6),(2.18) and (3.8)). $\underline{S}$ is chosen such that the expectation of Σ coincide with a fixed covariance matrix Σ for the Minnesota prior. As the mean and variance are given as[10]:

$$\mathbb{E}(\phi) = \underline{\phi} \quad \mathbb{V}\mathrm{ar}(\phi) = (\underline{\nu} - m - 1)^{-1}(\underline{S} \otimes \underline{\Omega}), \tag{2.27}$$

then the diagonal elements $\underline{S}$ are chosen in the following way:

$$(\underline{S})_{ii} = (\underline{\nu} - m - 1)\hat{\sigma}_i^2 \tag{2.28}$$

The choice of degrees of freedom of the inverse Wishart distribution $\underline{\nu}$ according to:

$$\underline{\nu} \geq \max\{m + 2, m + 2h - T\} \tag{2.29}$$

garantees the existence of both the prior variance of parameters and the posterior variance of forecasts at horizon h (see Kadiyala and Karlsson (1997)).

If the likelihood function (2.8) is taken into account, it is possible to show that the posterior distribution belongs to the same class (for example, Zellner (1996)):

$$\begin{cases} \Sigma|Y \sim \mathcal{IW}(\overline{S}, \overline{\nu}) \\ \Phi|\Sigma, Y \sim \mathcal{N}(\overline{\Phi}, \Sigma \otimes \overline{\Omega}) \end{cases} \tag{2.30}$$

[10]The RHS of 2.27 follows from: $\mathbb{V}\mathrm{ar}(\phi) = \mathbb{V}\mathrm{ar}(\mathbb{E}(\phi|\Sigma)) + \mathbb{E}(\mathbb{V}\mathrm{ar}(\phi|\Sigma)) = \mathbb{V}\mathrm{ar}(\underline{\phi}) + \mathbb{E}(\Sigma \otimes \underline{\Omega}) =$
$= \mathbb{E}(\Sigma \otimes \underline{\Omega}) = \mathbb{E}(\Sigma) \otimes \underline{\Omega} = (\nu - m - 1)^{-1}\underline{S} \otimes \underline{\Omega}$

with hyperparameters of the posterior distributions defined as:

$$
\begin{aligned}
&\overline{\nu} = \underline{\nu} + T \\
&\overline{\Omega} = (\underline{\Omega}^{-1} + X'X)^{-1} \\
&\overline{\Phi} = \overline{\Omega} \cdot (\underline{\Omega}^{-1}\underline{\Phi} + X'Y) \\
&\overline{S} = \underline{S} + \hat{E}'\hat{E} + \hat{\Phi}'X'X\hat{\Phi} + \underline{\Phi}'\underline{\Omega}^{-1}\underline{\Phi} - \overline{\Phi}'\overline{\Omega}^{-1}\overline{\Phi} \\
&\quad = \underline{S} + \hat{E}'\hat{E} + (\underline{\Phi} - \hat{\Phi})'(\underline{\Omega} + (X'X)^{-1})^{-1}(\underline{\Phi} - \hat{\Phi}), \text{ где:} \\
&\quad \hat{\Phi} = (X'X)^{-1}X'Y \text{ и } \hat{E} = Y - X\hat{\Phi}
\end{aligned}
$$

There is a popular alternative method to compute the parameters of the posterior distribution.

We set matrices $\underline{S}$ and $\underline{\Omega}^{-1}$ to zeros, consequently, the matrix $(\underline{\Omega} + (X'X)^{-1})^{-1}$ is also a zero matrix, and the matrix $\underline{\Phi}$ dissapears. To compensate the difference we add dummy observations to the matrices X and Y:

$$
X^* = \begin{bmatrix} X^+ \\ X \end{bmatrix}, \; Y^* = \begin{bmatrix} Y^+ \\ Y \end{bmatrix}
$$

When we add dummy observations, the matrices of scalar products $X^{*\prime}X^*$ and $X^{*\prime}Y^*$ are decomposed to a sum: $X^{*\prime}X^* = X^{+\prime}X^+ + X'X$, $X^{*\prime}Y^* = X^{+\prime}Y^+ + X'Y$. In a particular case, if dummy observations are zero, they do not change the matrices $X'X$, $X'Y$ and $Y'Y$ at all. Note, that the matrices X and Y are included in the hyperparameters of the posterior distribution as parts of the matrices $X'X$, $X'Y$ and $Y'Y$ only, and so the order of the dummy observations does not matter as well as how to add dummy observations with respect to the matrices X and Y. The dummy observations may be added in the beginning, in the end or even in the middle of the matrices X and Y.

We get new formulae for the posterior hyperparameters:

$$
\begin{aligned}
&\overline{\nu} = \underline{\nu} + T \\
&\overline{\Omega} = (X^{*\prime} X^{*})^{-1} = (X^{+\prime} X^{+} + X^{\prime} X)^{-1} \\
&\overline{\Phi} = \overline{\Omega} \cdot (X^{*\prime} Y^{*}) = \overline{\Omega} \cdot (X^{+\prime} Y^{+} + X^{\prime} Y) = (X^{*\prime} X^{*})^{-1} X^{*\prime} Y^{*} \\
&\overline{S} = \hat{E}^{*\prime} \hat{E}^{*} \\
&\hat{E}^{*} = Y^{*} - X^{*} \overline{\Phi}
\end{aligned}
$$

The observations are added so that the hyperparameters of the posterior distributions do not change. Therefore, it is necessary:

$$
\begin{cases}
X^{+\prime} X^{+} = \underline{\Omega}^{-1} \\
X^{+\prime} Y^{+} = \underline{\Omega}^{-1} \underline{\Phi} \\
(Y^{+} - X^{+} \underline{\Phi})^{\prime} (Y^{+} - X^{+} \underline{\Phi}) = \underline{S}
\end{cases}
\tag{2.31}
$$

The interdependence between the "new" formulae and the matrices determining the prior and posterior distributions are shown in a Table 2.1

	Interpretation	Formula
$\underline{\Phi}$	regression coefficient estimates Y^{+} to X^{+}	$(X^{+\prime} X^{+})^{-1} \cdot (X^{+\prime} Y^{+}) = \underline{\Phi}$
$\underline{S}$	scalar of residuals of these regressions	$\hat{E}^{+\prime} \hat{E}^{+}$, where $\hat{E}^{+} = Y^{+} - X^{+} \underline{\Phi}$
$\underline{\Omega}^{-1}$	scalar products of the regressors in X^{+}	$X^{+\prime} X^{+}$
$\overline{\Phi}$	coefficient estimates of the regressions Y^{*} to X^{*}	$(X^{*\prime} X^{*})^{-1} \cdot (X^{*\prime} Y^{*}) = \overline{\Phi}$
$\overline{S}$	scalar products of the residuals of the regressions	$\hat{E}^{*\prime} \hat{E}^{*}$, where $\hat{E}^{*} = Y^{*} - X^{*} \overline{\Phi}$
$\overline{\Omega}^{-1}$	scalar products of regressors in X^{*}	$X^{*\prime} X^{*}$

Table 2.1: 'Dummy observations and prior and posterior formulae'

These conditions are fulfilled if the dummy observations are added according to the scheme [11]:

[11] The similar formulae provided by Bańbura et al. (2010), Berg and Henzel (2013) to define the conjugate normal - inverse Wishart distributions are a particular case (3.12) for $\lambda_{lag} = 1$ и $\lambda_{const} \to \infty$.

$$
Y^{NIW} = \begin{bmatrix} \frac{\mathrm{diag}(\delta_1\sigma_1,\ldots,\delta_m\sigma_m)}{\lambda_{tight}} \\ 0_{m(p-1)\times m} \\ \mathrm{diag}(\sigma_1,\ldots,\sigma_m) \\ 0_{1\times m} \end{bmatrix} \quad X^{NIW} = \begin{bmatrix} \frac{\mathrm{diag}(1,2^{\lambda_{lag}},\ldots,p^{\lambda_{lag}})\otimes\mathrm{diag}(\sigma_1,\ldots,\sigma_m)}{\lambda_{tight}} & 0_{mp\times 1} \\ 0_{m\times mp} & 0_{m\times 1} \\ 0_{1\times mp} & \frac{1}{\lambda_{tight}\lambda_{const}} \end{bmatrix} \tag{2.32}
$$

A clear advantage of defining the conjugate normal - inverse Wishart distribution is that the prior is easily defined. Just several hyperparameters are enough to determine the prior. On the other hand, the moments of prior distributions of the parameters in different equations become interdependent. For example, all coefficients at the first lag of a dependent variable have a priori the same variance. Though this assumption usually is not very limiting but there are examples where the covariance matrix of the prior distribution cannot be formed symmetrically for different equations. The following example is widely known in the literature Kadiyala and Karlsson (1997)). Assume, a researcher wants to estimate a model and taking the neutrality of money into account. This assumption may be imposed with a prior where all coefficients at lags of money in the output equation have zero mean and low variance. However, it means that the prior variance of respective coefficients is also relatively low in other equations. This feature may be undesirable and, to avoid it, the researcher may opt for the independent normal - inverse Wishart prior distribution that we have already discussed above.

Modifications of the prior distribution

Doan et al. (1984) and Sims (1993) propose to take into account the characteristics of the actual time series that are not totally embedded in the common prior. This feature reflects the fact that many time series have unit roots and cointegration relations. The modification permits to avoid an incredibly large fraction of in-sample variance explained by exogenous variables. (Carriero, Clark, and Marcellino (2015)).

<u>Sum-of-coefficients prior</u>

A sum-of-coefficients prior was introduced by Doan et al. (1984). It expresses the common belief about the non-stationarity of the time series. If the variables in a VAR model have a unit root, then it is reasonable to impose a prior where a sum of all coefficients at lags of the dependent variable is equal to one. (Robertson and Tallman (1999) and Blake and Mumtaz (2012)). In other words, an average of lag values of a variable is a good forecast for future values of the variable.

This prior is imposed by adding dummy observations according to the following scheme.

$$Y^{SC} = \frac{1}{\lambda_{sc}} \left[\text{diag}(\delta_1 \mu_1, \ldots, \delta_m \mu_m) \right] \tag{2.33}$$

$$X^{SC} = \frac{1}{\lambda_{sc}} \left[(1_{1\times p}) \otimes \text{diag}(\delta_1 \mu_1, \ldots, \delta_m \mu_m) \quad 0_{m\times 1} \right], \tag{2.34}$$

with $(1_{1\times p})$ being a row vector of ones of length p, μ_i being an i-th component of the vector μ, which consists of average starting values of all variables [12]: $\mu = \frac{1}{p}\sum_{t=1}^{p} y_t$

Initial observation prior

The initial observation prior introduced by Sims (1993) reflects a prior belief that the variables have a common stochastic trend. It is imposed with just an only dummy observation so that all values of all variables are equal to the mean of starting values (for each variable, respectively) up to a scale coefficient λ_{io}:

$$Y^{IO} = \frac{1}{\lambda_{io}} \left[\delta_1 \mu_1, \ldots, \delta_m \mu_m \right] \tag{2.35}$$

$$X^{IO} = \frac{1}{\lambda_{io}} \left[(1_{1\times p}) \otimes (\delta_1 \mu_1, \ldots, \delta_m \mu_m) \quad 1 \right], \tag{2.36}$$

This prior implies that an average value of each variable is a linear combination of all other average values.

The hyperparameter λ_{io} is responsible for the tightness of this prior. When $\lambda_{io} \to 0$, all the variables are assumed to be either stationary with the mean equal to the sample average of starting values or nonstationary without drift and cointegrated.

As in case of Minnesota prior, the Gibbs algorithm is not needed, a random sample is generated explicitly from the posterior distribution. An example algorithm is as follows: :

- At iteration s generate:

$$\Sigma^{[s]} \sim \mathcal{IW}(\overline{S}, \overline{\nu})$$

$$\phi^{[s]} \sim \mathcal{N}(\overline{\phi}; \Sigma^{[s]} \otimes \overline{\Omega})$$

[12] Some athors take average of all observations in the sample: $\mu = \frac{1}{T}\sum_{t=1}^{T} y_t$ (Bańbura et al. (2010) and Carriero et al. (2015)). However, according to Sims and Zha (1998) the researcher is supposed to use first p observations only to calculate the average values.

In practice (as an option) instead of generating the vector $\phi^{[s]}$ it is possible to generate the matrix $\Phi^{[s]}$ in two steps:

1. Generate a matrix V of dimension $k \times m$ from independent standard normal random values.

2. Compute a matrix $\Phi^{[s]}$ according to the formula:[13]

$$\Phi^{[s]} = \overline{\Phi} + \text{chol}(\overline{\Omega}) \cdot V \cdot \text{chol}(\Sigma^{[s]})'$$

Therefore the raw observations are augmented with three blocks of dummy observations: Y^{NIW} and X^{NIW}, Y^{SC} and X^{SC}, Y^{IO} and X^{IO}. As besides Y^{NIW} and X^{NIW} two blocs of observations are also added, the structure of $\underline{\Omega}$ changes in comparison to (2.18) and (3.8).

2.2.3 Jeffreys prior distributions

The Jeffreys prior defines the prior distribution of the error covariance matrix without any hyperparametersbat all and it takes the form:

$$\Sigma \sim |\Sigma|^{-(m+1)/2} \tag{2.37}$$

In research papers the independent normal - Jeffreys prior and the conjugate normal - Jeffreys prior can be found.

Independent normal Jeffreys prior

$$\begin{cases} \phi \sim \mathcal{N}(\underline{\phi}; \Xi) \\ \Sigma \sim |\Sigma|^{-(m+1)/2} \\ \phi \text{ и } \Sigma \text{ independent} \end{cases} \tag{2.38}$$

[13]This formula is valid as the generation of the vector $\Phi \sim N(\bar{\Phi}, \Sigma \otimes \bar{\Omega})$: results from a Choleski decomposition of the covariance matrix $\Sigma \otimes \bar{\Omega}$: $\vec{\Phi} = \vec{\bar{\Phi}} + \text{chol}(\Sigma \otimes \bar{\Omega}) \times \nu$

This distribution is a special case of the independent normal - inverse Wishart if $\underline{S} = \underline{\nu}^{1/m} \cdot I$ and $\underline{\nu} \to 0$. The density function of the inverse Wishart distribution takes the form:

$$p(\Sigma) = \frac{1}{\Gamma_m(\underline{\nu}/2)} |\underline{S}|^{v/2} |\Sigma|^{-(\underline{\nu}+m+1)/2} 2^{-\underline{\nu}m/2} \operatorname{etr}\left(-\frac{1}{2}\underline{S}\Sigma^{-1}\right)$$

If $\underline{S} = \nu^{1/m} \cdot I$ и $\underline{\nu} \to 0$, then simultaneously $|\underline{S}|^{\nu} \to 1$ and $\operatorname{etr}\left(-\frac{1}{2}\underline{S}\Sigma^{-1}\right) \to 1$. So:

$$p(\Sigma) \to const \cdot |\Sigma|^{-(m+1)/2}$$

The Jeffreys distribution is improper, that means that it is impossible to scale the integral under the density function so that it is equal to unity. Nonetheless, the posteripr is proper if the number of observations is rather big. ($T > m - 1$) (Alvarez (2014)).

To get a sample from the posterior distribution the Gibbs algorithm can be used. The formulae for the hyperparameters of the posterior distributions can be derived from the general case by setting $\underline{S} = 0$, $\underline{\nu} = 0$.

The Minnesota prior and the independent normal - Jeffreys distributions are opposite special cases of the independent - inverse Wishart distribution. In case of the Minnesota distribution, the matrix Σ is assumed to be known and in case of the independent Jeffreys distribution the matrix Σ has a diffuse noninformative distribution.

Conjugate normal Jeffreys prior and diffuse Jeffreys prior

$$\begin{cases} \phi|\Sigma \sim \mathcal{N}(\underline{\phi}; \Sigma \otimes \underline{\Omega}) \\ \Sigma \sim |\Sigma|^{-(m+1)/2} \end{cases} \tag{2.39}$$

This distribution is a special case of the conjugate normal - inverse Wishart distribution if $\underline{S} = \underline{\nu}^{1/m} \cdot I$ и $\underline{\nu} \to 0$. The formulae for the posterior distributions can be obtained by setting: $\underline{\nu} = 0$, $\underline{S} = 0$.

A particular case for the two Jefferys priors is the diffuse Jefferys prior.

$$\begin{cases} \phi \sim 1 \\ \Sigma \sim |\Sigma|^{-(m+1)/2} \\ \phi \text{ и } \Sigma \text{ независимы} \end{cases} \tag{2.40}$$

To impose this prior no hyperparameter is needed. In can be derived from the independent normal - Jeffreys prior if $\underline{\phi} = 0$ and $\underline{\Xi} = a \cdot I$ и $a \to \infty$.

To obtain a sample from the posterior distribution it is possible to use a direct Monte-Carlo sampling without the Gibbs algorithm. The distribution can be derived from the general case by setting $\underline{S} = 0$, $\underline{\nu} = 0$, $\underline{\Xi}^{-1} = 0$, $\underline{\phi} = 0$. The formulae become substantially simpler, in particular there is no need any more to invert a matrix of the $km \times km$ dimension.

The diffuse Jefferys prior can also be obtained from the conjugate normal - Jeffreys prior distribution when $\underline{\phi} = 0$ and $\underline{\Omega} = a \cdot I$ and $a \to \infty$. The formulae for the hyperparameters of the posterior distribution can be obtained by setting $\underline{\phi} = 0$, $\underline{\Omega}^{-1} = 0$, $\underline{\nu} = 0$, $\underline{S} = 0$.

The fact that Jefferys distributions depend on a very few hyperparameters can be considered both as an advantage and as a drawback of Jeffreys distributions. On one hand, the researcher has to think less about the choice of hyperparameters. On the other hand, a small number of hyperparameters leads to the nonflexibility of these prior distributions.

2.2.4 Interdependence of prior distributions

All interdependencies between the prior distributions discussed so far in the paper are shown in the graph.

$$\begin{tikzcd}
Minnesota \arrow[rr, "\underline{\Xi}=\Sigma\otimes\underline{\Omega}"] & & Minnesota-\otimes \\
iN-IW \arrow[u, "\underline{S}=(\underline{v}-m-1)\Sigma,\ \underline{v}\to\infty"] \arrow[d, "\underline{S}=\underline{v}^{1/m}I,\ \underline{v}\to 0"'] & & cN-IW \arrow[u, "\underline{S}=(\underline{v}-m-1)\Sigma,\ \underline{v}\to\infty"'] \arrow[d, "\underline{S}=\underline{v}^{1/m}I,\ \underline{v}\to 0"] \\
iN-J \arrow[r, "\underline{\phi}=0,\ \underline{\Omega}=aI,\ a\to\infty"] & D-J & cN-J \arrow[l, "\underline{\phi}=0,\ \underline{\Omega}=aI,\ a\to\infty"']
\end{tikzcd}$$

$Minnesota$ is the Minnesota prior

$Minnesota-\otimes$ is the Minnesota prior with a Kronecker-type covariance matrix $\underline{\Xi} = \Sigma \otimes \underline{\Omega}$

$iN - IW$ is independent normal - inverse Wishart prior

$cN - IW$ is conjugate normal - inverse Wishart prior

$iN - J$ is independent normal - Jeffereys prior

$cN - J$ is conjugate normal - Jeffereys prior

$D - J$ is noninformative - Jeffereys prior

The arrows go from more general distributions to special cases, the restrictions are given above the arrows.

2.2.5 A data-driven choice of hyperparameters: large BVAR estimation

Optimal choice of the tightness hyperparameter

In some cases the prior hyperparameters are defined endogenously. Particularly, it happens when a model of a large dimension (with many series) is estimated. It was shown that large BVAR models can show better forecasting performance than small ones (however it is not always the case). Though small-size BVAR have been widely used since beginning of 1990s when there were introduced in the literature, the usage of large-sample BVAR models was extremely limited till recently. An opinion shared by many researcher stated that the Bayesian shrinkage itself is not enough for solving the over-parametrization problem, and so it is necessary to impose some additional non-Bayesian restrictions anyway.

The papers by De Mol, Giannone, and Reichlin (2008) and Bańbura et al. (2010), played the key role in the evolution of the new approach. They showed that BVAR can be estimated on large sample without additional non-Bayesian restrictions. However, large samples require that the parameter λ_{tight} increases if the dimension of the dataset increases, it means that the prior distribution is tighter for larger samples. In the literature two algorithms to define the optimal λ_{tight} in this framework can be found.

A first algorithm was introduced by Bańbura et al. (2010) and it implies that the shrinkage is so tight to avoid the over-parametrization. It also assumes that a three-variable VAR model does not contain too many parameters and does not require any additional shrinkage. It means that the parameter λ_{tight} can be chosen in such a way that a large BVAR model has the same fit as a three-variable VAR. In other words, each model is shrunk to the size of an unrestricted three-variable VAR.

Denote an actual value of a variable var at the moment $T+h$ as $y_{var,T+h}$, and a forecast of the variable var, made at the moment T for a horizon h using a model with m variables and the tightness parameter λ as: $y^{\lambda,m}_{var,T+h|T}$. The algorithm of choosing λ consists of following steps:

Step 1. Compute an in-sample one-period forecasts on a training sample and calculate the mean squared forecast error (MSFE) for M variables in a set ($\mathcal{M}$) that the researcher is the most interested in.

$$MSFE^{\lambda,m}_{var,1} = \frac{1}{T_0 - p} \sum_{t=p}^{T_0-1} \left(y^{\lambda,m}_{var,t+1|t} - y_{var,t+1} \right)^2 , \tag{2.41}$$

with T_0 being the last observation in the training sample.

Step 2. Compute one-period forecasts for the random-walk model with drift [14] for the same variables $\left(MSFE^0_{var,1}\right)$ and calculate a new indicator $FIT^{\lambda,m}$ reflecting an average relative MSFE:

$$FIT^{\lambda,m} = \frac{1}{M} \sum_{var \in \mathcal{M}} \frac{MSFE^{\lambda,m}_{var,1}}{MSFE^0_{var,1}} \tag{2.42}$$

Step 3. Estimate a three-variable VAR for the same M variables that are the most interesting for the forecaster [15] and we compute MSFE and indicator $FIT^{\infty,M}$:

$$FIT^{\infty,M} = \frac{1}{M} \sum_{var \in \mathcal{M}} \frac{MSFE^{\infty,M}_{var,1}}{MSFE^0_{var,1}} \tag{2.43}$$

Step 4. To find numerically the optimal value of λ take the value that minimize the difference between $FIT^{\lambda,m}$ and $FIT^{\infty,M}$:

$$\lambda^*_m = \arg\min_{\lambda} |FIT^{\lambda,m} - FIT^{\infty,M}| \tag{2.44}$$

Afrer the optimal λ is chosen for each model,an out-of-sample forecast can be done on a testing sample.

[14]MSFE for BVAR and VAR are normalized for MSFE, obtained with the RW model to take into account the fact different unis of measurement for different series. We use a subscript 0 for forecasts obtained with the RW model as the RW model is a particular case for the BVAR for $\lambda = 0$ and $\delta_i = 1, i = 1, \ldots, k$.

[15]We use a subscript ∞ for forecasts obtained with an unrestricted VAR model as the unrestricted VAR is a special case for the BVAR model if $\lambda \to \infty$. In this case the posterior coincides with the likelihood function.

One-period forecast shrinkage

Another algorithm is proposed by Doan et al. (1984) and requires to choose the optimal parameter λ_{tight} so that to maximize the accuracy of the one-period out-of-sample forecast on the training sample. The procedure amounts to the maximization of the marginal density function:

$$\lambda^* = \arg\max_{\lambda} \ln p(Y) \tag{2.45}$$

The marginal density function can be obtained by integration of the model coefficients:

$$p(Y) = \int p(Y|\phi)p(\phi)d\phi \tag{2.46}$$

If the prior is conjugate normal - inverse Wishart distribution, the marginal density can be computed analytically (Zellner (1996); Bauwens, Lubrano, and Richard (2000); Carriero, Kapetanios, and Marcellino (2012)):

$$\begin{aligned} p(Y) = \pi^{-\frac{Tm}{2}} \times + \left|(I + X\underline{\Omega}X')^{-1}\right|^{\frac{N}{2}} \times |\underline{S}|^{\frac{\underline{\nu}}{2}} \times \frac{\Gamma_N(\frac{\underline{\nu}+T}{2})}{\Gamma_N(\frac{\underline{\nu}}{2})} \times \\ \times \left|\underline{S} + (Y - X\underline{\Phi})'(I + X\underline{\Omega}X')^{-1}(Y - X\underline{\Phi})\right|^{-\frac{\underline{\nu}+T}{2}}, \end{aligned} \tag{2.47}$$

where $\Gamma_N(\cdot)$ is a N-dimensional gamma function. The choice of the number of lags is done in a similar way by maximization of the marginal density on p (2.47):

$$p^* = \arg\max_{p} \ln p(Y) \tag{2.48}$$

Some aspects of coding

The Gibbs or Monte Carlo algorithms often require inverting positive-definite symmetrical matrices. Some of those matrices have a determinant close to zero that precludes the matrix inverse on a computer. In this case a researcher can use a following method

1. Do a Cholesky decomposition for a given matrix A

   $$A = U'U,$$

   where U is an upper triangular matrix

2. Inverse matrix U. There are special algorithms to inverse upper triangular matrices.

3. Compute A^{-1} following the formula:

$$A^{-1} = U^{-1}U^{-1\prime}$$

However even this method may imply technical difficulties if the matrix is numerically degenerate. In this case it is possible to use a Moore-Penrose pseudo-inverse matrix.

If it is also known that the matrix A can be written as $A = X'X$, then it is possible to compute an inverse matrix A^{-1} without computing A:

1. To do a singular decomposition of the matrix X, $X = U\Sigma V'$

2. To compute an inverse matrix to $A = X'X$ using a formula $A^{-1} = V\Sigma^{-2}V'$

2.3 BVAR forecasting

2.3.1 Posterior forecasting density

Bayesian VAR models in reduced form are usually estimated for forecasting purposes. BVAR permit to do both point and density forecasts. If an evaluation of the quality of the forecast is needed as well, then the model is estimated on a historical sample, and the forecasts are done for those periods for which actual values are already available. A model can be estimated either on a sliding (rolling) window or on an expanding window (recursive regression). In the former case the estimation is done on the same number of observations but the start and the end of the sample shift for an observation at each step. The forecasts are done for a selected forecasting horizon on each step as well. The process goes till all observations that permit to compare forecasts and actual data are used up. In case of expanding window, the start of a sample is fixed and the length of the sample increases by one observation at each step.

A key concept for making forecasts with a Bayesian model is posterior predictive density function, which we denote as $p(y_{T+1:T+H}|Y_T)$ following Karlsson (2013). The notation means that a forecast is done for all time periods starting at $T+1$ and ending as $T+H$ if the actual values till the moment T are available. Here the matrix $y_{T+1:T+H} = (y_{T+1} \ldots, y_{T+H})'$ shows future observations for this period and the matrix $Y_T = (y_1 \ldots, y_T)'$ shows all observations used for the model estimation. The posterior predictive density function can be written as:

$$p(y_{T+1:T+H}|Y_T) = \int p(y_{T+1:T+H}|Y_T, \phi)p(\phi|Y_T)d\phi, \tag{2.49}$$

where $p(y_{T+1:T+H}|Y_T, \phi)$ is the density function of future observations given parameters ϕ and data till the period T, and $p(\phi|Y_T)$ denote the posterior density function of parameters.

In general, an analytical expression for the predictive density is not available for a forecasting horizon longer than 1 period. For longer horizons, the predictive density is computed numerically according to the formula (2.49). To do it, the forecasting values of the variables $\tilde{y}_{T+h}$ at a horizon $h = 1, \ldots, H$ are computed using the conditional density function $p(y_{T+1:T+H}|Y_T, \phi)$ for each realization of parameters from posterior distribution $p(\phi|Y_T)$. In case of making a forecast for a forecasting horizon h, all the forecasted values for an horizon $\tilde{h} < h$ are assumed to be known. If the procedure is repeated many times, the researcher obtains a sample from the posterior predictive distribution for each h.

In other words, for the BVAR model the posterior predictive density is computed according the following scheme (Karlsson (2013), p. 800, 811):

1. Generate sets of parameters from the posterior distribution. For Minnesota prior, conjugate normal - inverse Wishart prior and independent normal - inverse Wishart prior the algorithms laid out on pages 68, 75, and 70, respectively, can be used.

2. At iteration s generate $\varepsilon^{[s]}_{T+1}, \ldots, \varepsilon^{[s]}_{T+H}$ from $\varepsilon_t \sim N(0, \Sigma^{[s]})$ (in case of Minnesota prior $\Sigma^{[s]} = \Sigma$) and compute recursively:

$$\tilde{y}^{[s]}_{T+h} = \Phi^{[s]}_{ex} + \sum_{i=1}^{h-1} \Phi^{[s]}_i \tilde{y}^{[s]}_{T+h-i} + \sum_{i=h}^{p} \Phi^{[s]}_i y^{[s]}_{T+h-i} + \varepsilon^{[s]}_{T+h} \tag{2.50}$$

It is necessary to ignore that different forecasts are computed for different values of parameters. For Minnesota prior and conjugate normal-inverse Wishart prior $\{\tilde{y}^{[s]}_{T+1}, \ldots, \tilde{y}^{[s]}_{T+H}\}^S_{s=1}$ are considered as a sample of independent draws from a joint predictive distribution.In case of independent normal-inverse Wishart distribution $\varepsilon^{[s]}$ is generated only if s is greater than a given B, as first B draws of posterior parameters are used for the chain convergence (this is a burn-in period) and dropped from the entire sample. Therefore, the forecast draws are also taken into account only for $s > B$: $\{\tilde{y}_{T+1}^{[s]}, \ldots, \tilde{y}_{T+H}^{[s]}\}_{s=B+1}^{S}$.

In case of conjugate normal - inverse Wishart prior, the process of parameter generation can be accelerated. To do it the parameter matrix Φ can be computed as:

$$\Phi = \bar{\Phi} + \mathrm{chol}(\bar{\Omega}) \times V \times \mathrm{chol}(\Sigma)', \tag{2.51}$$

A one-step ahead forecast is a linear function of parameters and therefore the posterior predictive density for $h = 1$ can be derived analytically. It takes the form of a matrix t-distribution with parameters depending on the prior distribution used for the estimation. In case of conjugate normal - inverse Wishart distribution (Carriero et al. (2015) p. 54) the forecast has a multivariate t-distribution with parameters:

$$y'_{T+1}|x'_{T+1} \sim MT\left(x'_{T+1}\bar{\Phi}, \left(x'_{T+1}\bar{\Omega}x_{T+1}\right)^{-1}, \bar{S}, \bar{\nu}\right) \tag{2.52}$$

A forecast for a horizon longer than one period requires a numerical procedure laid out above.

Point forecasts are computed using a random sample from from posterior forecasting distribution. A choice of a type of the point forecast (for example, a mode or a median of the forecasting density) can be done using a loss function. From a formal point of view, a researcher has a loss function $\mathbb{L}(a, y_{T+1:T+H})$, that determines which value matrix a is chosen as a point forecast. The value matrix is chosen so that the expected losses given available data Y_T are minimized (Karlsson (2013), page. 795):

$$\mathbb{E}[\mathbb{L}(a, y_{T+1:T+H})|Y_T] = \int \mathbb{L}(a, y_{T+1:T+H})p(y_{T+1:T+H}|Y_T)dy_{T+1:T+H} \tag{2.53}$$

Given a loss function and the predictive density, the solution of the minimization problem is a function of available data only $a(Y_T)$. For particular cases of a loss function the solution takes a simple form. For example, for a quadratic loss function:

$$\mathbb{L}(a, y_{T+1:T+H}) = (a - y_{T+1:T+H})'(a - y_{T+1:T+H})$$

the solution takes a form of the conditional mean, $a(Y_T) = \mathbb{E}(y_{T+1:T+H}|Y_T)$, and for the loss function expressed as an absolute value, the solution is the median of the predictive distribution.

In applied papers on BVAR both point and density forecasts can be found. The BVAR models are often compared with alternative forecasting models in terms of their forecasting accuracy.

2.3.2 The evaluation of a point forecast accuracy: a univariate case

To evaluate the accuracy of a point forecast, the mean squared forecasting error (MSFE) and root mean squared forecasting error are widely used in applied papers.

$$MSFE^M_{var,h} = \frac{1}{N_h}\sum_T (y^M_{var,T+h|T} - y_{var,T+h|T})^2, \tag{2.54}$$

$$RMSFE^M_{var,h} = \sqrt{MSFE^M_{var,h}}, \tag{2.55}$$

with $y^M_{var,T+h|T}$ being a forecast for the variable var, made at the moment T for h steps ahead with a model M. A number of these forecasts made at different T for the forecasting horizon h is denoted by N_h.

An alternative measure of the forecasting accuracy is the mean absolute forecast error:

$$MAFE^M_{var,h} = \frac{1}{N_h}\sum_T |y^M_{var,T+h|T} - y_{var,T+h|T}| \tag{2.56}$$

2.3.3 The evaluation of a point forecast accuracy: a multivariate case

In a multivariate case, when a researcher has to evaluate the forecasting accuracy for several variables using just one indicator, two statistics are in use, both suggested by Adolfson, Lindé, and Villani (2007). There are a trace and a log-determinant of the matrix of mean squared errors $\Sigma_A(h)$ that is calculated as:

$$\Sigma_A(h) = \frac{1}{N_h}\sum_{t=T}^{T+N_h-1} \tilde{\varepsilon}_{t+h|t}\tilde{\varepsilon}'_{t+h|t}, \tag{2.57}$$

with $\tilde{\varepsilon}_{t+h|t} = A^{-1/2}\varepsilon_{t+h|t}$, a $\varepsilon_{t+h|t}$ being a forecasting error at a horizon h and A is an arbitrary positive definite matrix. The ranking of the forecasts does not depend on the choice of the matrix A if the log-determinant is used as the indicator and does depend if the trace is used, as $\ln|\Sigma_A(h)| = \ln|\Sigma_I(h)| - \ln|A|$, and $\operatorname{tr}|\Sigma_A(h)| = \operatorname{tr}|A^{-1}\Sigma_I(h)|$.

For example, to choose a matrix A, Adolfson et al. (2007) take a diagonal matrix of sample variances of forecasted variables. In this case the trace of the matrix is equal to the weighted average of forecast mean squared errors for individual time series. However, Adolfson et al. (2007) claim that point forecast ranking made with multivariate indicators may be misleading. The reason is that the statistics are highly influenced by inaccurately forecasted variable. Such variables usually contain a significant part of the last main component but these variables are not necessary interesting for the researcher. (Adolfson et al. (2007)).

2.3.4 The comparison of point forecast accuracy

In empirical application, a researcher usually has a choice of models to be used for forecasting of one of several variables. Therefore, a question rises if the models have the equal forecast accuracy. The most widely used test of equal forecast accuracy of two models is Diebold-Mariano test (Diebold and Mariano (1995)). Let e_{1t} and e_{1t} be the vectors of forecast errors obtained with two competing models: $e_{Mt} = y^M_{var,t|t-h} - y_{var,t}, M = 1, 2$ and $g(e_{Mt})$ is a loss function associated with the forecast errors [16]. Then we can define $d_t = g(e_{1t}) - g(e_{2t}), t = 1, \ldots, n$ as loss differential and $\bar{d}_t = \frac{1}{n}\sum_{t=1}^{n} d_t$ as sample mean of loss differential and $\mu = E(d_t)$ as population mean of the loss differential. It is possible to show that under H_0 of equal forecast accuracy:

$$\sqrt{(}T)(\bar{d} - \mu) \to N(0, 2\pi f_d(0)), \tag{2.58}$$

where $f_d(0)$ is the spectral density of the loss differential at frequency 0: $f_d(0) = \frac{1}{2\pi}\sum_{k=-\infty}^{k=\infty} \gamma_d(k)$ where $\gamma_d(k)$ is autocovariance of the loss function at lag k. In practice, a convenient estimator of $2\pi f_d(0)$ is $\sum_{k=-W}^{k=W} \hat{\gamma}_d(k)$, where $W = T^{\frac{1}{3}}$. Therefore the Diebold-Mariano statistics takes the form:

$$DM = \frac{\bar{d}}{\sqrt{\frac{\sum_{k=-W}^{k=W} \hat{\gamma}_d(k)}{T}}} \tag{2.59}$$

[16]For example, it might be squared error loss $g(e^2_{Mt})$ or absolute error loss $|g(e^2_{Mt})|$.

which has N(0,1) asymptotic distribution under zero hypothesis of equal forecast accuracy of the two models in question. As the Diebold-Mariano test can be over-sized in small samples especially for longer forecast horizons, Harvey, Leybourne, and Newbold (1997) propose a small sample correction of the Diebold-Mariano statistics:

$$HLN = \sqrt{\frac{n + 1 - 2h + n^{-1}h(h-1)}{n}} DM \tag{2.60}$$

The authors also show that comparing the statistics with critical values from the Student's t distribution with $n - 1$ degrees of freedom is more appropriate than with those from normal distribution.

The extension of the Diebold-Mariano test that permits to compare the forecast accuracy for several models was proposed by Mariano and Preve (2012). A zero hypothesis states that all models in a available set - containing M models - have equal forecasting accuracy:

$$E(d_t) = 0, \tag{2.61}$$

where $d_t = (d_{1t}, \ldots, d_{M-1,t})'$ and $d_{jt} = g(e_{j,t} - e_{j+1,t}), j = 1, \ldots, M-1$. Mariano and Preve (2012) show that under weak condition $n\bar{d}'\bar{\Omega}^{-1}\bar{d}$ is asymptotically distributed as χ^2_{M-1} under H_0, where $\bar{\Omega}$ is a consistent estimator of the asymptotic variance $\Omega = \Gamma(0) + \sum_{h=1}^{q}(\Gamma(h) + \Gamma'(h))$ and $Gamma(h)$ is the autocovariance matrix of d_t at lag h. The finite sample correction proposed by Harvey et al. (1997) for DM statistics can also be applied in case of multivariate Diebold-Mariano test. Whereas Diebold-Mariano's statistics tests for equal predictive accuracy, there are papers that propose tests for superior predictive ability (for example, White (2000), Hansen (2005), Romano and Wolf (2005)). Tests for equal predictive ability make a ground for more elaborate procedures such as model confidence set procedure proposed by Hansen, Lunde, and Nason (2011) that is applied in the third chapter of this dissertation and discussed there in details.

2.3.5 The evaluation of a density forecast accuracy

Posterior density function

A point forecast provides a probable future value of a variable but does not provide any information about the uncertainty of the forecast. A density forecast clearly shows the uncertainty of forecast. A density forecast can be demonstrated with fan charts similar to those that some central banks use to present their inflation forecast. Contrary to ordinary confidence intervals, a density forecast is not symmetrical relative to the point forecast in general. A principal difficulty for a density forecast evaluation is that the true density is not observable, and only are actual values of forecast variables observable, just one observation for a time period.

In the literature "the statistical consistency between the distributional forecasts and the observations" is called calibration Gneiting, Balabdaoui, and Raftery (2007). Gneiting et al. (2007) and Mitchell and Wallis (2011) evaluate calibration accuracy in their papers. A researcher can have different attitude toward possible discrepancies between the forecast and an actual value. A widely used method to calibrate a loss function is called a probability integral transform, or PIT.[17] According to this method, the actual values are substituted into the posterior cumulative distribution function(Diebold, Gunther, and Tay (1998), Gerdrup, Jore, Smith, and Thorsrud (2009), Gonzalez-Rivera and Sun (2015)).

Assume that a forecast at time τ has the cumulative distribution function denoted by $F(\cdot|Y_T)$, and the actual value of a variable being forecast is denoted by y_τ. If the observed values are actually a sample from this distribution, than the function values $F(y_\tau|Y_T)$ have a uniform distribution on $[0, 1]$. Forecast quality is evaluated with a histogram. If an actual cumulative distribution is the same as assumed, the columns of the histogram should have approximately the same height.

Contrary to the predictive density analysis, the PIT method is a frequentist one as it compares the distribution of actual values with a potential one which they would have if the data generation process corresponds to the model. Some formal tests based on PIT can be found in Geweke and Amisano (2010).

We can underline two problems linked to PIT application. First, it is difficult to verify that PIT have really uniform distribution. Even if a model in hand is a DGP for a variable, the predictive density may not show it due the uncertainty of parameters. Second, Gneiting et al. (2007) show using simulations that it is impossible to choose the true model anambiguously among several models based on PIT. As an alternative Gneiting et al.

[17]The name refers to the fact that that cumulative distribution function is comuted by taking the integral of the density function.

(2007) propose to maximize "the sharpness of the predictive density given the calibration". The sharpness of calibration refers to the concentration of distributions around actual realizations that can be evaluated with box plots or scoring rules.

Scoring rules

Scoring rules are functions according to which a forecast receives a certain number of scoring points (scores) depending the actual value of the forecasted variable. A detailed review of scoring rules is presented by Tsyplakov (2013). If a researcher compares several forecasting models, then the model with maximal number of scores is considered as the most accurate.

A scoring rule widely applied in academic literature. A rule most commonly used is log predictive density scores suggested by Good (1952) and described by Geweke and Amisano (2010). Adolfson et al. (2007), Christoffel, Warne, and Coenen (2010) and Carriero et al. (2015) show some recent example of the application of this rule.

The score is calculated as follows:

$$s_h = \sum_{t=T}^{T+N_h-1} \ln p(y_{t+h}|Y_t) \tag{2.62}$$

The score for a one-step forecast can be expressed using the marginal likelihood function (Adolfson et al. (2007), p. 324-325):

$$\begin{aligned} s_1 = \ln[p(y_{T+1}|Y_T) \cdot \ldots \cdot p(y_{T+N_1}|Y_{T+N_1-1})] = \\ = \ln[p(y_{T+1}, \ldots, y_{T+N_1}|Y_T)] = \ln m(T+N_1) - \ln m(T), \end{aligned} \tag{2.63}$$

with $m(t) = p(y_1, \ldots, y_t) = \int p(y_1, \ldots, y_t|\phi)p(\phi)d\phi$ being the marginal likelihood function of all data till the moment t, and $p(\phi)$ being a prior density. Under the integral, there are only the prior density function $p(\phi)$ and the likelihood function $p(y_1, \ldots, y_t|\phi)$. Therefore, when the marginal density is computed, the actual data is not used. It permits to interpret the marginal likelihood as a measure of the accuracy of the out-of-sample forecast and not as a measure of the in-sample fit.

A computation of s_h for the horizon $h > 1$ is more complicated as the density function $p(y_{t+h}|Y_t)$ does not have an explicit analytic expression. One of the approaches consists in the density function estimation $p(y_{t+h}|Y_t)$ based on realizations of the forecasts with a kernel density estimator. In practical terms this method can be applied only if the number of variables in the model is not large. In case of many variables Adolfson et al. (2007) propose to assume that $p(y_{t+h}|Y_t)$ is the multivariate normal density and to estimate a mean vector and its covariance matrix on a forecast sample.

For example, Carriero et al. (2015) take the assumption of the normal distribution and apply the following algorithm to calculate the score:

1. Generate a sample of forecasts for the variable var at a horizon h
2. Compute the estimate of the logarithm of forecast density according to the formula:

$$
\begin{aligned}
s_t\left(y_{t+h}^{var}\right) &= \log p(y_{t+h}|Y_t, m) = \\
&= -0.5\left[\ln(2\pi) + \ln(V_{t+h|t}^{var}) + (y_{t+h}^{var} - \bar{y}_{t+h|t})^2 / V_{t+h|t}^{var}\right]
\end{aligned}
\tag{2.64}
$$

   with $p(y_{t+h}|Y_t, m)$ being a marginal prediction conditional density for $y_{t+h}, \dots, y_{t+h}$, depending on observable data $Y_t = \{y_1, \dots, y_T\}$, and the vector $\bar{y}_{t+h|t}^t$ and matrix $V_{t+h|t}^{var}$ denote the posterior mean and the variance of forecast distribution for the variable var and forecasting horizon h.

3. Compute a score as a mean of the estimates for each forecasting horizon: $s_t\left(y_{t+h}^{var}\right)$:

$$
\bar{s}_{var,h}^M = \frac{1}{N_h} \sum s_t\left(y_{t+h}^{var}\right), \tag{2.65}
$$

   with N_h being a number of forecasts for a horizon h.

To provide the statistical significance of the differences in average log scores a researcher can use a t-test of equal means suggested by Amisano and Giacomini (2007) that can be applied to the log score of each model in the available set of competing models relative to the baseline forecast.

2.4 Conclusion

This paper surveys the techniques for estimation and forecasting with the reduced-form BVAR models. We explain in detail different prior distributions widely used in macroeconomics applied papers and we make 'a map' for them that contains a detailed description of their interdependence. A separate section of the paper is devoted to the algorithm of defining the conjugate normal - inverted Wishart distribution with dummy observations. This method is widely used in applied papers but is not described in many existing surveys of BVARs. In the section devoted to forecasting we consider both point and density forecasts.

Appendices

.1 Available code realizations

Source	Software	Min	Conj N-IW	Ind N-IW	SoC	IO
Carriero	Matlab	-	?	-	+	+
Blake Mumtaz	Matlab	-	+	+	+	+
Koop Korobilis	Matlab	+	+	+	-	-
Zha	Matlab		?		+	+
Le Sage	Matlab	?				
Sims	Matlab		?	?	+	+
Canova	Matlab		?			
BMR	R	+	-	+	-	-
MSBVAR	R	-	+	-	+	+
bvarr	R	+	+	+	+	+
Sims	R		?	?	+	+
Built-in function	EViews	+	+	-	+	+
Built-in function	Dynare	?	+	?	+	+

1. Carriero: Dummy observations are used as for cNIW prior but at the same time the Gibbs sampler is implemented, $\overline{\Phi}$ is fixed and Σ is recomputed at each iteration depending on the previous Φ. If the matrix $X'X$ is badly scaled, then a pseudo inverse matrix is used. There is a part of the code that generates VAR coefficients if the eigenvalues are outside of the unit circle, and this chunk is not explained in the code. http://cremfi.econ.qmul.ac.uk/efp/info.php

2. Blake Mumtaz: The Independent NIW is also called Minnesota prior. A code for conjugate NIW is written in the same way as Carriero did using Gibbs Sampling. There is a slight difference in the code, as Blake and Mumtaz add two identical rows for dummy when the prior variance for a constant is defined. http://www.bankofengland.co.uk/education/Pages/ccbs/technical_handbooks/techbook4.aspx

3. Koop Korobilis: The code is not flexible. The code should be modified to make forecasts for a horizon longer than one period. The baseline cNIW and iNIW priors do not contain hyperparameters and are defined with fixed matrices. https:

//sites.google.com/site/dimitriskorobilis/matlab

4. Zha: According to their paper, the restrictions are imposed on the structural form of the VAR model. Therefore, some hyperparameters are interpreted in a differernt way.

5. Sims: The description is not detailed. It is necessary to read all the code to modify something. http://sims.princeton.edu/yftp/VARtools/

6. BMR: The simulations are realized in C++. The package is also suitable to estimate DSGE and TVP-BVAR mddels. The package description is good. http://bayes.squarespace.com/bmr/

7. MSBVAR: The simulations are realized in Fortran and C++. The package is also suitable to estimate Markov-switching BVARs. https://cran.r-project.org/web/packages/MSBVAR/

8. bvarr: The code for cNIW prior is flexible. If the matrix $X'X$ is badly scaled, then a pseudo-inverse matrix is used. The code for Minnesota and iNIW prior is the translation of the code by Koop and Korobilis. This part of the code is less flexible. https://github.com/bdemeshev/bvarr

9. bvarsv: TVP, https://github.com/FK83/bvarsv

10. Eviews: The code ignores the fact that the coefficients are estimated by the Bayesian methods, the forecasts are made exactly as in a frequentist model The coefficient λ_{kron} is equal 0,99 and cannot be changed. The prior means for all first lags coefficients must be identical for all variables.

11. Dynare: The function is presented as à la Sims. The estimation is only possible in a package but a user can change a prior for the covariance matrix. http://www.dynare.org/

.2 Table of notations

Notation	Dimension	Description	Formula
p	scalar	number of lags	
m	scalar	number of endogenous variables	
d	scalar	number of exogenous variables	
k	scalar	number of parameters in one equation	$k = mp + d$
T	scalar	number of observations	
z_t	$d \times 1$	vector of exogenous variables (including a constant)	
y_t	$m \times 1$	vector of endogenous variables	$y_t = \Phi' x_t + \varepsilon_t$
x_t	$k \times 1$	vector of all regressors	$x_t = [y'_{t-1} \dots y'_{t-p} \ z'_t]'$
ε_t	$m \times 1$	vector of random errors	$y_t = \Phi' x_t + \varepsilon_t$
Y	$T \times m$	all endogenous variables	$Y = [y_1, y_2, \dots, y_T]'$
X	$T \times k$	matrix of regressors	$X = [x_1, x_2, \dots, x_T]'$
E	$T \times m$	matrix of errors	$E = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T]'$
y	$mT \times 1$	vectorization of Y	$y = \vec{(}Y)$
ε	$mT \times 1$	vectorization of E	$\varepsilon = \vec{(}E)$
$\Phi_1, \dots$	$m \times m$	VAR coefficients	$y_t = \Phi_1 y_{t-1} + \dots + \Phi_{const} + \varepsilon_t$
Φ_{const}	$m \times d$	vector of all constants	$y_t = \Phi_1 y_{t-1} + \dots + \Phi_{const} + \varepsilon_t$
Φ	$k \times m$	matrix combination $\Phi_1, \dots$	$\Phi = [\Phi_1 \dots \Phi_p \ \Phi_{const}]'$
ϕ	$km \times 1$	vector of matrices Φ	$\phi = \vec{\Phi}$

$\underline{\Phi}$	$k \times m$	prior expected value Φ	
$\underline{\phi}$	$km \times 1$	vector from a matrix $\underline{\Phi}$	$\underline{\phi} = \vec{\underline{\Phi}}$
$\overline{\Phi}$	$k \times m$	posterior expected value Φ	
$\overline{\phi}$	$km \times 1$	vector from a matrix $\overline{\Phi}$	$\overline{\phi} = \vec{\overline{\Phi}}$
$\underline{\Xi}$	$km \times km$	prior covariance matrix Φ	
$\overline{\Xi}$	$km \times km$	posterior covariance matrix Φ	$\overline{\Xi} = (\underline{\Xi}^{-1} + \Sigma^{-1} \otimes X'X)^{-1}$
$\underline{\nu}$	scalar	prior number of degrees of freedom	
$\overline{\nu}$	scalar	posterior number of degreees of freedom	$\overline{\nu} = T + \underline{\nu}$
$\underline{\Omega}$	$k \times k$	matrix of prior scaling coefficients of the covariance matrix Φ	$\underline{\Xi} = \Sigma \otimes \underline{\Omega}$
$\overline{\Omega}$	$k \times k$	Matrix of posterior scaling coefficients of the covariance matrix Φ	$\overline{\Omega} = (\underline{\Omega}^{-1} + X'X)^{-1}$, $\overline{\Xi} = \Sigma \otimes \overline{\Omega}$
Σ	$m \times m$	Covariance matrix of errors	$\mathbb{E}\,\varepsilon_t \varepsilon_t' = \Sigma$

.3 Correspondence of hyperparameters in different papers

DM18	CCM15	BGR10, BH13	KK97
λ_{tight}	λ_1	λ	$\sqrt{\pi_1}$
λ_{kron}	$\lambda_2 = 1$	$\vartheta = 1$	$\sqrt{\pi_2/\pi_1}$
λ_{lag}	1	1	0.5
λ_{const}	λ_0/λ_1	∞	$\sqrt{\pi_3/\pi_1}$
λ_{exo}	NA	NA	NA
λ_{sc}	λ_3	τ	
λ_{io}	λ_4	NA	

DM18 this paper, CCM15 Carriero et al., 2015, BGR10 Bańbura et al., 2010, BH13 Berg and Henzel, 2013, KK97 Kadiyala and Karlsson, 1997

Chapter 3

Forecasting Russian macroeconomic indicators with BVAR[1]

3.1 Introduction

Accurate macroeconomic forecasts are extremely important for policy making. Central banks and government bodies monitor a large set of macroeconomic indicators to determine the policy (Beckner (1996), Bernanke and Boivin (2003)). Therefore, a model used for forecasting must be suitable for data-rich samples because large models might outperform low-dimensional ones by taking into account more potentially relevant information. This explains the recent resurgence in interest from academics, central bankers and private sector experts for macroeconomic forecasting in a data-rich environment.

In this paper, we forecast Russian macroeconomic indicators with Bayesian vector autoregressions (BVARs) of different sizes. Our goal is twofold. First, we compare the forecast accuracy of BVAR with that of unrestricted vector autoregressions (VARs) and random walk with drift models for 23 important macroeconomic indicators. Second, we question whether a high-dimensional model always outperforms a low-dimensional one in terms of forecasting accuracy.

For the last 30 years, VARs introduced by Sims (1980) have become a widely-used tool for forecasting. However, unrestricted VARs bear the risk of over-parametrization even for samples of moderate size. This risk stems from the fact that the number of parameters

[1]co-authored with Boris Demeshev, NRU HSE

to be estimated increases nonlinearly with the number of equations. For this reason, in economic applications unrestricted VARs usually contain only up to eight variables, and this may potentially lead to the loss of some relevant information and undermine the forecast accuracy.

To deal with a data-rich environment researchers modify VARs and impose restrictions on the covariance structure. One strand of the literature focuses on dynamic factor models (DFM, Forni, Hallin, Lippi, and Reichlin (2000) and Stock and Watson (2002))and Panel VARs and Global VARs (PVARs, GVARs, Pesaran, Schuermann, and Weiner (2004) and Dees and Güntner (2014)). DFM are based on the idea that a relatively small set of indices extracted from a high-dimensional set of variables can summarize the information from this set. These factors are treated as variables in a VAR model either separately or in conjunction with several time series from the original information set in a factor-augmented VAR (FAVAR) model. For data sets with a panel structure a suitable choice is a PVAR or a GVAR with shrinkage done by exclusion, exogeneity or homogeneity restrictions.

Another method of shrinkage is the Bayesian one and we follow this approach. The shrinkage is done by imposing restrictions on the parameters in the form of prior distributions. While BVARs in a low-dimensional space were widely used for macroeconomic analysis, their use for data-rich environments was limited until recently. The reason was a general agreement that Bayesian shrinkage is insufficient to solve the over-parametrization problem in high cross-sectional dimension samples.

However, in their influential paper, De Mol et al. (2008) show that Bayesian methods can be successfully applied to a data-rich environment if the degree of shrinkage is set relative to the cross-sectional dimension of the sample. Bańbura et al. (2010) confirm and develop this assertion for BVARs applied to a large set of US time-series. Their main result is that high-dimensional models have better forecasting performance than small-dimensional models and even FAVARs. They also show that accurate forecasts can be already obtained using a medium-sized BVAR (20 variables in their case).

Several authors have recently shown that, in terms of forecasting accuracy, medium and large BVARs outperform their low-dimensional counterparts. For example, Beauchemin and Zaman (2011) present a medium BVAR with a good forecasting performance applied to the US data. Bloor and Matheson (2010) compare univariate autoregresions (ARs), unrestricted VARs and BVARs and show evidence that high-dimensional BVARs, in general demonstrate better forecasting performance. Koop (2013) demonstrates that high-dimensional BVARs outperform factor models in terms of forecasting performance. Moreover, he argues that more complicated priors than those that are usually applied may not lead to more precise forecasts. Alessandri and Mumtaz (2014) underline the importance of financial factors for an accurate forecast of output and inflation, especially «for predicting «tail» macroeconomic outcomes». Carriero et al. (2015) study some characteristics of BVARs and find those providing the most accurate forecasts.

Our analysis delivers two important results. First, we show that most Russian macroeconomic indicators in our sample can be forecast by BVARs more accurately than by competing models. However, contrary to other studies (for example, Bloor and Matheson (2010), Bańbura et al. (2010)) we do not confirm that relative forecast error monotonically decreases with the dimension of the sample. In almost half of those cases where a BVAR is the most accurate model, a small-dimensional BVAR outperforms its high-dimensional counterpart.

The paper is structured as follows. Section 2 presents our model and the prior distribution we apply. In Section 3 we describe our sample and the data transformations we use. Section 4 contains the results and their interpretation. Section 5 concludes.

3.2 Model

3.2.1 BVAR

Let y_{it} be variables[2] stacked in a $m \times 1$ vector $y_t = (y_{1t}, y_{2t}, \ldots, y_{mt})'$. The reduced form VAR can be written as:

$$y_t = \Phi_{const} + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \ldots + \Phi_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma) \tag{3.1}$$

where $\Phi_{const} = (c_1, \ldots, c_m)'$ is a $m \times 1$ vector of constants, Φ_l are autoregression $m \times m$ - dimensional matrices where $l = 1, \ldots, p$. Vector ε_t is a m-dimensional vector of errors with covariance matrix $\mathbb{E}\,\varepsilon_t \varepsilon_t' = \Sigma$, and is uncorrelated with regressors. By grouping

[2]For the convenience of the reader, all the notations are also shown in Appendix 1.

parameter matrices into one matrix $\Phi = [\Phi_1 \dots \Phi_p \;\; \Phi_{const}]'$ and defining new vector $x_t = [y'_{t-1} \;\dots\; y'_{t-p} \; 1]'$, the equation (3.1) can be written in a more compact form:

$$y_t = \Phi' x_t + \varepsilon_t \tag{3.2}$$

If the variables and shocks are grouped in the following way: $Y = [y_1, y_2, \dots, y_T]'$, $X = [x_1, x_2, \dots, x_T]'$, $E = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T]'$, the VAR can be written as:

$$Y = X\Phi + E \tag{3.3}$$

The Bayesian estimate combines a likelihood function $L(Y|\Phi, \Sigma)$ with a prior distribution $p(\Phi, \Sigma)$ and results in a posterior distribution of parameters $p(\Phi, \Sigma|Y)$:

$$p(\Phi, \Sigma|Y) \propto p(\Phi, \Sigma) L(Y|\Phi, \Sigma) \tag{3.4}$$

3.2.2 Conjugate normal — inverse Wishart prior

Our benchmark model for estimation and forecasting purposes is a BVAR with a conjugate normal — inverse Wishart prior. The prior can be written as:

$$\begin{cases} \Sigma \sim \mathcal{IW}(\underline{S}, \underline{\nu}) \\ \Phi|\Sigma \sim \mathcal{N}(\underline{\Phi}, \Sigma \otimes \underline{\Omega}) \end{cases} \tag{3.5}$$

The prior mean of the coefficient matrices is written with a $k \times m$ matrix $\underline{\Phi} = \mathbb{E}(\Phi)$, where $\underline{\Phi} = [\underline{\Phi}_1 \dots \underline{\Phi}_p \; \underline{\Phi}_{const}]'$.The matrices $\underline{\Phi}_l$ are defined as follows:

$$(\underline{\Phi}_l)_{ij} = \begin{cases} \delta_i \; i = j, l = 1; \\ 0, \quad \text{otherwise} \end{cases} \tag{3.6}$$

A matrix $\underline{\Omega}$ is assumed to be diagonal and it depends on several hyperparameters:

$$\underline{\Omega} = \mathrm{diag}\{\underline{\Omega}_{lag=1}, \dots, \underline{\Omega}_{lag=p}, \underline{\Omega}_{const}\} \tag{3.7}$$

$$(\underline{\Omega}_{lag=l})_{jj} = \left(\frac{\lambda}{l^{\lambda_{lag}} \hat{\sigma}_j}\right)^2 \quad \underline{\Omega}_{const} = \lambda^2_{const} \tag{3.8}$$

The hyperparameters have the following interpretation: λ determines the overall tightness of the prior and it is responsible for the relative weight of the prior with respect to the information incorporated in the data, λ_{lag} controls the velocity of the decrease of the prior variance with increasing the lag length, and λ_{const} governs the relative tightness of the prior for the constant terms.

The scale matrix $\underline{S}$ is diagonal and its non-zero elements assure that the mean of Σ is equal to the fixed covariance matrix of the standard Minnesota prior:

$$(\underline{S})_{ii} = (\underline{\nu} - m - 1)\hat{\sigma}_i^2 \tag{3.9}$$

The scale parameter σ_i^2 is usually set to be equal to the variance estimate of residuals in a univariate AR model. The choice of degrees of freedom of inverse Wishart distribution $\underline{\nu}$ greater than or equal to than $\max\{m+2, m+2h-T\}$ guarantees the existence of the prior variance of the regression parameters and the posterior variances of the forecasts at horizon h (Kadiyala and Karlsson (1997)).

It is possible to show that the posterior distribution formed by combining this prior distribution with a likelihood function is also normal — inverse Wishart (see, for example, Zellner (1996)):

$$\begin{cases} \Sigma|Y \sim \mathcal{IW}(\overline{S}, \overline{\nu}) \\ \Phi|\Sigma, Y \sim \mathcal{N}(\overline{\Phi}, \Sigma \otimes \overline{\Omega}) \end{cases} \tag{3.10}$$

with the following parameters:

$$\begin{aligned}
&\overline{\nu} = \underline{\nu} + T \\
&\overline{\Omega} = (\underline{\Omega}^{-1} + X'X)^{-1} \\
&\overline{\Phi} = \overline{\Omega} \cdot (\underline{\Omega}^{-1}\underline{\Phi} + X'Y) \\
&\overline{S} = \underline{S} + \hat{E}'\hat{E} + \hat{\Phi}'X'X\hat{\Phi} \\
&\quad + \underline{\Phi}'\underline{\Omega}^{-1}\underline{\Phi} - \overline{\Phi}'\overline{\Omega}^{-1}\overline{\Phi} \\
&\hat{\Phi} = (X'X)^{-1}X'Y \\
&\hat{E} = Y - X\hat{\Phi}
\end{aligned}$$

There is a popular alternative approach to calculate hyperparameters of the posterior distribution. We set $\underline{S}$ and $\underline{\Omega}^{-1}$ to be zero matrices and to compensate the difference we add supplementary observations into X and Y matrices according to:

$$Y^* = \begin{bmatrix} Y^{NIW} \\ Y \end{bmatrix} \quad X^* = \begin{bmatrix} X^{NIW} \\ X \end{bmatrix}, \tag{3.11}$$

where matrices Y^{NIW} and X^{NIW} are defined as follows [3]:

$$Y^{NIW} = \begin{bmatrix} \frac{\text{diag}(\delta_1\sigma_1,\ldots,\delta_m\sigma_m)}{\lambda} \\ 0_{m(p-1)\times m} \\ \text{diag}(\sigma_1,\ldots,\sigma_m) \\ 0_{1\times m} \end{bmatrix} \quad X^{NIW} = \begin{bmatrix} \frac{\text{diag}(1,2^{\lambda_{lag}},\ldots,p^{\lambda_{lag}})\otimes\text{diag}(\sigma_1,\ldots,\sigma_m)}{\lambda} & 0_{mp\times 1} \\ 0_{m\times mp} & 0_{m\times 1} \\ 0_{1\times mp} & \frac{1}{\lambda_{const}} \end{bmatrix} \tag{3.12}$$

This method permits the calculation $\overline{\Phi}$ as an OLS estimate of the regression of Y^* on X^*: $\overline{\Phi} = (X^{*\prime}X^*)^{-1}X^{*\prime}Y^*$ and $\overline{S}$ as a sum of the squared residuals for this regression: $\overline{S} = (Y^* - \overline{\Phi}X^*)'(Y^* - \overline{\Phi}X^*)$.

3.2.3 Prior modifications

Doan et al. (1984) and Sims (1993) propose complementing this prior distribution with additional information in form of two other priors. This modification reflects the belief that time series may have unit roots and cointegration relations. These elements in the prior allow avoiding an unreasonably large share of the variation in the data which is accounted for by deterministic components (Sims (1993)).

A sum-of-coefficients prior was introduced by Doan et al. (1984). If all the time-series in a sample have a unit root, this information can be taken into account with a prior where a sum of all the lag parameters for each dependent variable is equal to one (Robertson and Tallman (1999), Blake and Mumtaz (2012)). In other words, when the mean of the lagged values of a variable is at a certain level, this level is a good forecast for future observations of this dependent variable. We implement this prior by combining the dataset given in 3.11 with artificial dummy-observations according to the following scheme:

[3]The similar formulae provided in Bańbura et al. (2010), Berg and Henzel (2013) can be regarded as special cases of (3.12) for $\lambda_{lag} = 1$ and $\lambda_{const} \to \infty$.

$$Y^{SC} = \frac{1}{\lambda_{sc}} \Big[\mathrm{diag}(\delta_1\mu_1, \ldots, \delta_m\mu_m)\Big] \tag{3.13}$$

$$X^{SC} = \frac{1}{\lambda_{sc}} \Big[(1_{1\times p}) \otimes \mathrm{diag}(\delta_1\mu_1, \ldots, \delta_m\mu_m) \quad 0_{m\times 1}\Big], \tag{3.14}$$

where $(1_{1\times p})$ is a unitary $[1 \times p]$ vector, μ_i is i-th component of vector μ, which contains the average values of initial observations of all variables in the sample[4]: $\mu = \frac{1}{p}\sum_{t=1}^{p} y_t$.

The dummy initial observation prior proposed by Sims (1993) expresses the belief that the variables have a common stochastic trend. Only one observation is added so that the values of all variables are equal to the average value of initial observations μ_i normalized to a scale coefficient λ_{io}. Therefore, this extra observation is defined as follows:

$$Y^{IO} = \frac{1}{\lambda_{io}} \Big[\delta_1\mu_1, \ldots, \delta_m\mu_m\Big] \tag{3.15}$$

$$X^{IO} = \frac{1}{\lambda_{io}} \Big[(1_{1\times p}) \otimes (\delta_1\mu_1, \ldots, \delta_m\mu_m) \quad 1\Big], \tag{3.16}$$

This prior distribution reflects the belief that the average value for a variable is a linear combination of average values of all the other variables.

The hyperparameter λ_{io} controls the tightness of this prior. When $\lambda_{io} \to 0$, the model implies that either all variables are stationary with the mean equal to sample mean of the initial observations or non-stationary without drift and cointegrated.

3.2.4 Choice of tightness hyperparameter: the algorithm of shrinkage

As shown by De Mol et al. (2008) and confirmed in several other recent studies, a sample with a larger cross-sectional dimension requires a lower λ, so the prior must be tighter for a larger sample than for a smaller one. In this paper, we use the approach introduced by Bańbura et al. (2010) to determine the optimal λ for every model.

This algorithm is based on the idea that the shrinkage should be sufficiently tight to avoid over-parametrization. Moreover, it is assumed that a three-variable unrestricted

[4]Some authors calculate μ using the average values of all observations in a sample, so that $\mu = \frac{1}{T}\sum_{t=1}^{T} y_t$ (Bańbura et al. (2010) and Carriero et al. (2015)). However, following Sims and Zha (1998) we calculate μ using only initial p observations.

VAR is parsimonious enough and does not require any additional shrinkage. This implies that the hyperparameter λ can be chosen so that the model has the same in-sample fit as a three-variable VAR. In other words, a BVAR model of any dimension is shrunk to the size of a small unrestricted VAR. A detailed description of the procedure is laid out below. We denote the actual value of a variable var at moment $T+h$ by $y_{var,T+h}$, and a forecast of the variable var at moment T for a horizon h in a model with m variables and an overall tightness parameter λ by $y^{\lambda,m}_{var,T+h|T}$. The algorithm for choosing λ has the following steps.

1. We make in-sample one-period forecasts with BVAR on a training sample and calculate the mean squared forecast error for the set of M variables of central interest[5]:

$$MSFE^{\lambda,m}_{var,1} = \frac{1}{T_0 - p} \sum_{t=p}^{T_0-1} \left(y^{\lambda,m}_{var,t+1|t} - y_{var,t+1} \right)^2, \tag{3.17}$$

   where the BVAR coefficients are obtained using the training sample: $t = p+1, \dots, T_0$ and T_0 is the last observation of the training sample: $T_0 = p + 120$.

2. In a similar way we calculate one-period forecasts according to the random walk model[6] for the same variables $\left(MSFE^{0}_{var,1}\right)$ and a new indicator $FIT^{\lambda,m}$:

$$FIT^{\lambda,m} = \frac{1}{M} \sum_{var \in \mathcal{M}} \frac{MSFE^{\lambda,m}_{var,1}}{MSFE^{0}_{var,1}} \tag{3.18}$$

3. We estimate VARs for the same set of M variables of interest[7] and calculate MSFEs and an indicator $FIT^{\infty,M}$:

$$FIT^{\infty,M} = \frac{1}{M} \sum_{var \in \mathcal{M}} \frac{MSFE^{\infty,M}_{var,1}}{MSFE^{0}_{var,1}} \tag{3.19}$$

[5]Our benchmark set of variables of central interest ($\mathcal{M}$) includes the industrial production index, consumer price index and interbank interest rate so that $M = 3$. As a robustness check we excluded the interest rate from this set and there was almost no change in the vector of the optimal λ.

[6]We normalize MSFE for the BVAR and VAR models by MSFE obtained with the random walk model to take into account the different scales of the series. We use a superscript 0 for the random walk model as random walk may be considered as a special case of BVAR if $\lambda = 0$ and $\delta_i = 1, i = 1, \dots, k$.

[7]We denote all results from VAR by a superscript ∞ as unrestricted VAR is a special case of BVAR with $\lambda \to \infty$. In this case the posterior coincides with the likelihood function.

4. The optimal lambda is the value minimizing the difference between $FIT^{\lambda,m}$ and $FIT^{\infty,M}$:

$$\lambda^*_m = \arg\min_{\lambda} |FIT^{\lambda,m} - FIT^{\infty,M}| \tag{3.20}$$

After the optimal λ is chosen for every m, we keep it fixed and make out-of-sample forecasts on the evaluation sample.

3.2.5 Out-of-sample forecasting

We estimate BVARs with the optimal λ on «rolling window» containing 120 observations, starting from observation $p+1$ and continuing until March 2015. The first p observations are used as a pre-sample and the subsample $[p+1, p+120]$ is a training sample to determine the optimal λ on a grid. We denote the last available observation as T_1, and the last observation of each evaluation subsample as τ. The number of forecasts is equal to $T_1 - T_0 - h + 1$ where h is the forecasting horizon ($h = 1, 3, 6, 9, 12$). Therefore, the number of one-period forecasts is greater than the number of three-period forecasts by two, etc.[8] For every model m and forecasting horizon h we calculate the out-of-sample MSFE for all m variables included in the model:

$$OMSFE^{\lambda,m}_{var,h} = \frac{1}{T_1 - T_0 - h + 1} \sum_{\tau=T_0}^{T_1-h} \left(y^{\lambda,m}_{var,\tau+h|\tau} - y_{var,\tau+h} \right)^2, \tag{3.21}$$

Then we calculate the MSFE of out-of-sample forecasts obtained with random walk with drift ($OMSFE^{0}_{var,h}$) and unrestricted VAR ($OMSFE^{\infty,m}_{var,h}$):

$$OMSFE^{\infty,m}_{var,h} = \frac{1}{T_1 - T_0 - h + 1} \sum_{\tau=T_0}^{T_1-h} \left(y^{\infty,m}_{var,\tau+h|\tau} - y_{var,\tau+h} \right)^2 \tag{3.22}$$

$$OMSFE^{0}_{var,h} = \frac{1}{T_1 - T_0 - h + 1} \sum_{\tau=T_0}^{T_1-h} \left(y^{0}_{var,\tau+h|\tau} - y_{var,\tau+h} \right)^2, \tag{3.23}$$

To compare the forecast accuracy of different models we report the relative MSFE, that is, the ratio of the MSFE of the model in question by the MSFE of a reference model (random walk in our case):

[8] An alternative method is to calculate an equal number of forecasts for each horizon h, starting from $T_0 + 12$. However this implies the loss of some information about the forecasts and we do not proceed with this method here.

$$RMSFE_{var} = \frac{OMSFE^{\lambda,m}_{var,h}}{OMSFE^{0}_{var,h}} \tag{3.24}$$

3.3 Data and Estimations

Our dataset consists of 23 time series running from January 1996 to April 2015. Our sample containing 232 observations is limited by data availability. The full list of the series and their sources is displayed in Appendix 2. We seasonally adjust data which demonstrate seasonal fluctuations with TRAMO/SEATS option in EViews and apply logarithms to the series, with the exception of those already expressed in rates.

We estimate models of different cross-sectional dimension. The industrial production index, CPI and interbank interest rate are forecast with three-variable, six-variable and 23-variable models. Monetary aggregate, the real effective exchange rate and the oil price index are forecast with four-variable, six-variable and 23-variable models. All the other series are forecast with four-variable, seven-variable and 23-variable models. For all models with dimension less than eight we estimate both unrestricted VARs and BVARs. We estimate only a BVAR on the sample with 23 variables. The three-variable VAR is the simplest specification that can be justified by a textbook version of a New Keynesian model. A model with six variables is specified in line with many monetary models used previously for the structural analysis of different economies (Sims (1992), Kim and Roubini (2000), Bjørnland (2008), Scholl and Uhlig (2008)) and it contains the real effective exchange rate, monetary aggregate M2, and the oil price index in addition to three variables included in the smallest VAR. The oil price index is used as a variable in the model to reflect the belief that oil price index is an important explanatory factor for the other variables in the sample as Russia has a petroleum export-based economy. To forecast variables outside of these core sets we estimate four-variable and seven-variable VARs containing three or six-variable samples described above plus an additional variable of interest. We include all available time series in our 23-variable model. In totally, after the optimal λ is chosen, we estimate 79 models. In a compact form the models used for forecasting are presented in Table 3.1. For VARs and BVARs we take all possible lags from 1 to 12.

Table 3.1: List of models and variable sets

VAR3/BVAR3	$Y = \{IP, CPI, R\}$
VAR4/BVAR4	$Y = \{IP, CPI, R, Z\}$
VAR6/BVAR6	$Y = \{IP, CPI, R, M2, REER, OPI\}$
VAR7/BVAR7	$Y = \{IP, CPI, R, M2, REER, OPI, W\}$
BVAR23	Y includes all 23 variables from the dataset

where IP is the industrial production index, CPI is the consumer price index, R is the nominal interbank rate, $M2$ is the monetary aggregate M2, $REER$ is the real effective exchange rate, OPI is the Brent oil price index. Z is any variable from the dataset besides IP, CPI and R. W is any variable from the dataset besides IP, CPI, R, $M2$,$REER$, and OPI.

3.4 Results

For every variable and every forecasting horizon we find a model with the lowest RMSFE. We compare 60 specifications for each variable and each forecasting horizon as we have 5 models (a VAR and a BVAR with 3 or 4 variables, a VAR and a BVAR with 6 or 7 variables, and a BVAR with 23 variables) and 12 lags for each of them. We visualize our results with color tables (Figures 3.1-3.2). The two tables in these Figures differ by the hyperparameter sets used for the BVAR priors. For models depicted in Figure 3.1 we take $\delta_i = 1$ for nonstationary series and $\delta_i = 0.5$ for stationary series while constructing the prior. We use the KPSS test (Kwiatkowski, Phillips, Schmidt, and Shin (1992) to split the series into two groups. The parameters σ_i are taken to be equal to the standard deviations of the residuals in the univariate $AR(p)$ model. This hyperparapeter set will be referenced as set A in what follows. Figure 3.2 is related to BVARs with the prior determined by the univariate $AR(1)$ model. We take δ_i as equal to the OLS estimates of the first lag parameter and σ_i as equal to the standard deviations of residuals in AR(1) models. This set will be referenced as set B.

The color of a cell corresponds to the model that appears to outperform the others in

	h=1	h=3	h=6	h=9	h=12
ind product	0.92		0.96	0.82	0.7
cpi	0.44	0.38	0.46	0.38	0.33
interb rate			0.66	0.52	0.58
agriculture	0.93	0.82	0.7	0.67	0.57
construction	0.97				
employment	0.67	0.42	0.43	0.6	0.72
export	0.59	0.61	0.76	0.81	0.89
gas price	0.73	0.43	0.22	0.29	0.5
gov balance	0.61	0.79	0.77	0.7	0.63
import	0.75	0.48	0.52	0.72	0.98
labor request	0.66	0.79	0.94	0.95	0.96
lend rate	0.94	0.84	0.77	0.77	0.8
M2	0.53	0.51	0.71	0.95	
nominal ER					
NFA of CB	0.6	0.56	0.75	0.65	0.6
oil price	0.88	0.81	0.88	0.81	0.77
ppi	0.43	0.75	0.69	0.59	0.49
real income	0.87	0.84	0.83	0.71	0.73
real invest	0.78	0.61	0.73	0.88	0.91
real ER	0.72	0.68	0.6		
retail	0.62	0.39	0.4	0.64	0.88
unemp rate	0.93	0.83	0.9	0.92	0.94
wage	0.74	0.51	0.46	0.42	0.41

Random walk

VAR3/4 - BVAR 3/4

VAR6/7

BVAR6/7

BVAR23

Figure 3.1: RMSFE of the best forecasting accuracy models, parameter set A: σ_i are std of $AR(p)$ residuals, $\delta_i = 1$ for nonstationary series, $\delta_i = 0.5$ for stationary series

terms of forecasting accuracy for a given variable and a given forecasting horizon. Most of cells are green (either light green or bright green) reflecting that a BVAR provides the most accurate forecast for the corresponding variables and forecasting horizons. An unrestricted VAR gives the most accurate forecast for variables and horizons indicated by yellow and orange cells. The procedure for choosing λ is such that the BVAR and the unrestricted VAR necessarily coincide for the smallest sample (3 or 4 variables). This explains why orange represents both of these models. Blue means that neither BVAR nor VAR beat the random walk in terms of forecast accuracy.

The forecast accuracy is measured with RMSFE calculated according to (3.24) and is also shown in Figures 3.1-3.2. The numbers less than one indicate that the a VAR or a BVAR model provides a better forecast than the random walk and the smaller the number is, the more accurate the forecasts are relative to the random walk. We see that in most cases we have at least one model that provides a forecast much better than the reference model.

	h=1	h=3	h=6	h=9	h=12
ind product	0.96		0.96	0.82	0.7
cpi	0.38	0.37	0.46	0.36	0.27
interb rate			0.91	0.56	0.56
agriculture	0.93	0.82	0.7	0.67	0.56
construction	0.97				
employment	0.7	0.54	0.59	0.7	0.81
export	0.57	0.62	0.71	0.8	0.86
gas price	0.7	0.42	0.22	0.31	0.51
gov balance	0.6	0.79	0.78	0.74	0.64
import	0.74	0.63	0.82	0.88	0.97
labor request	0.66	0.79	0.94	0.94	0.95
lend rate	0.95	0.89	0.79	0.71	0.66
M2	0.55	0.6	0.8	0.97	
nominal ER					
NFA of CB	0.6	0.62	0.69	0.61	0.61
oil price	0.85	0.81	0.85	0.79	0.75
ppi	0.43	0.74	0.69	0.6	0.49
real income	0.91	0.93	0.84	0.73	0.75
real invest	0.81	0.63	0.76	0.88	0.92
real ER	0.72	0.69	0.8		
retail	0.62	0.4	0.45	0.72	0.88
unemp rate	0.94	0.91	0.89	0.9	0.92
wage	0.75	0.53	0.47	0.42	0.41

- Random walk
- VAR3/4 - BVAR 3/4
- VAR6/7
- BVAR6/7
- BVAR23

Figure 3.2: RMSFE of the best forecasting accuracy models, parameter set B: σ_i are std of $AR(1)$ residuals, $\delta_i = 1$ are first lag $AR(1)$ estimates

Despite different prior parameter sets, the two tables are very similar both in terms of the best forecasting models and the relative accuracy with respect to the random walk.

We interpret our results as follows. First, for many variables and forecasting horizons, BVARs outperforms the random walk and unrestricted VARs. Out of the 115 forecasting cases highlighted in the paper (23 variables times 5 forecasting horizons) BVARs appear to be best in terms of forecast accuracy in 71 cases for the prior hyperparameter set A and in 77 cases for the prior hyperparameter set B. There are variables in our sample that are forecast more accurately by BVARs for all forecasting horizons we try (such as employment, import and lending rate). For several variables BVARs are the best option for the shortest horizons (for example, monetary aggregate M2 and the real effective exchange rate). On the contrary, for the agricultural production index a BVAR model has the lowest forecast error only for a one-year horizon.

Second, among all cases where BVARs show their forecasting accuracy, a high-dimensional BVAR is the best option in about half of the cases (35 of 71 for set A and 39 of 77 for set B). In other cases it is beaten by a low-dimensional BVAR (6 or 7 variables).

Third, for some variables and some forecasting horizons neither unrestricted VARs nor BVARs outperform the random walk. For example, in all specifications we consider the nominal exchange rate cannot be forecast by either VARs or BVARs better than by the random walk, and it is a long-held consensus in economics remounting to Meese and Rogoff (1983). However, we question another wide-spread belief that the price of oil is a random walk process. We show that the oil price index can be forecasted by BVARs much better than by the random walk and the result is robust for different prior settings.

3.5 Robustness check: the MCS algorithm

The method of choosing the best model laid out in the previous section has an evident drawback. We choose just an only one model for each variable and each forecasting horizon - called «the first best» - with the lowest RMSE whereas the difference between «the first best» and «the second best» (or even some nth best for $n > 2$) models might be insignificantly low. To check this option we employ the model confidence set procedure (MCS) proposed by Hansen et al. (2011). This procedure encompasses the series of tests that results in a confidence set, that is a set of models for which the hypothesis of equal predictive ability cannot be rejected at a certain confidence level. Let M be an initial set of models of dimension m, and each model gives the certain loss (a function of a prediction error) for each observation that we denote by l. We use a quadratic loss function:

$$l^i_{var,t} = (y^i_{var,t|t-h} - y_{var,t})^2 \quad i = 1, \dots, m, \tag{3.25}$$

where $y^i_{var,t|t-h}$ is an h-step forecast of a variable var made for the period t with a model i; $y_{var,t}$ is an observed value of the variable, and $l^i_{var,t}$ is a loss given for a variable var for the period t with a model i.

Loss differential, denoted by $d_{ij,t}$, is calculated for all pairs of models:

$$d_{ij,t} = l_{i,t} - l_{j,t}, \quad i, j = 1, \dots, m \tag{3.26}$$

The average loss of a model i with respect to a particular model j is calculated as:

$$d_{ij} = \frac{1}{T}\sum_{t} d_{ij,t} \quad i = 1, \ldots, m, \quad j \neq i \tag{3.27}$$

The average loss of a model i with respect to all other competitor models is calculated also for all i:

$$d_i = \frac{1}{m-1}\sum_{j \in M} d_{ij} \quad i = 1, \ldots, m, \quad j \neq i \tag{3.28}$$

Two intermediate statistics, t_i and t_{ij} are calculated in the following way:

$$t_i = \frac{d_i}{se(d_i)} \qquad t_{ij} = \frac{d_{ij}}{se(d_{ij})}, \tag{3.29}$$

where se is the standard error estimate calculated with a block-bootstrap procedure. The statistics t_i is interpreted as an indicator of the average inferiority of a model with respect to all the other models in question. The statistic t_{ij} is interpreted as an indicator of the inferiority of a model with respect to a specific model j.

As discussed in Hansen et al. (2011), the statistics used for testing the hypothesis of the equal predictive ability are calculated as:

$$T_R = max_{i,j \in M}|t_{ij}| \qquad T_{max} = max_{i \in M} t_i \tag{3.30}$$

The statistic T_R is an indicator of the difference between the losses of the most and the least accurate models, and T_{max} is an indicator of the difference between the loss of the least accurate model and the loss of other competing models on average. Both statistics have non-standard distributions. As it was said before, the MCS procedure goes as follows. First of all, a hypothesis about equal forecasting accuracy of all models in the set is tested. If the hypothesis is rejected than the worst model is eliminated from the set and the hypothesis is tested again for a smaller set. As a result, implementing the MCS procedure delivers a set of superior models with equal (insignificantly different) predictive accuracy.

Figure 3.3 represents the confidence sets for five principal variables and five forecast horizons. We demonstrate the number of specifications that are included in the confidence set if we rely on T_{max} statistics.[9] For every VAR or BVAR model the maximum number

[9]The results for all 23 variables as well as the results obtained when using T_R statistics are available upon request.

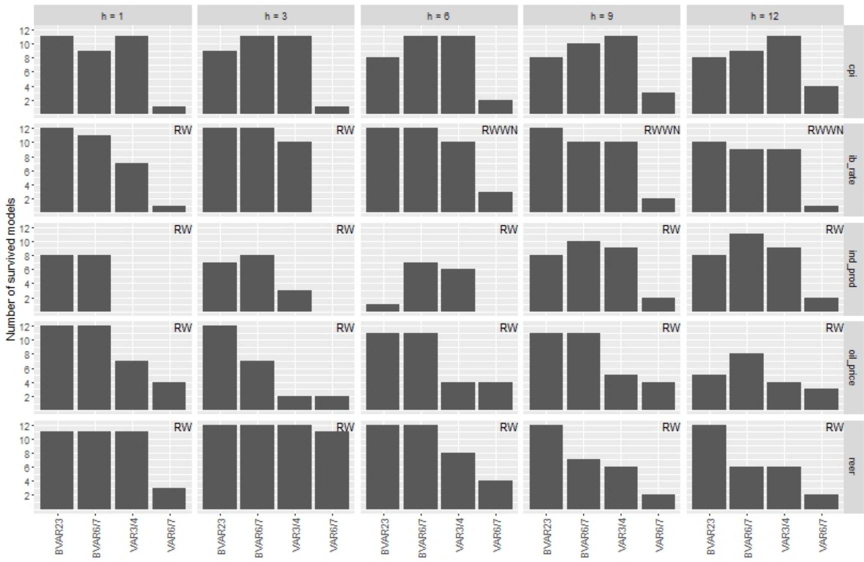


Figure 3.3: Robustness check with MCS algorithm

of specifications that might be included in the set is equal to 12 as we estimate models with all possible lags up to 12. Forecasts made with BVAR3/4 models necessarily coincide by construction with those made with VAR3/4 so we do not display the same columns twice. The figure shows that the BVAR columns are regularly higher than the VAR columns. It means that the confidence set encompasses BVAR models more often than VAR models. This confirms our previous claim that on a general basis the BVARs demonstrate better forecasting performance than VARs, though there are usually at least several VAR specifications whose forecasting performance is not significantly inferior to that of BVARs. However, it is necessary to mention that in most cases the RW is included in the confidence set as well. Therefore, even if a BVAR specification gives a more accurate forecast in terms of RMSFE, the difference in accuracy of this specification with the simple RW is not significantly different. Among five principal variables shown in the Figure3.3 the only exception is the price index. For this variable the RW model is not included in the confidence set. This variable is also an exception in another aspect: the VAR3/4 model with all possible lags under consideration is included in the confidence set for all forecasting horizons. For interbank interest rate, even the WN model is included in the confidence set at longer forecasting horizons.

3.6 Robustness check 2: Alternative benchmark models

In the previous discussion we compared the forecast accuracy of competing models with random walk. However, the random walk forecast might be less accurate than a forecast obtained with another simple model. To challenge this issue, we estimated two univariate models - ARIMA and ETS - on all 23 variables in the sample. The parameters of ARIMA were chosen with an automatic algorithm suggested in *forecast* package in R (Hyndman (2019)). First, the automatic algorithm finds the order of integration of the series and takes differences to make the series stationary. Then the model is selected among a set ARMA models (ARMA(0,0), ARMA(1,0), ARMA(0,1), ARMA(2,2)) Finally, the autoregression and moving average orders are increased and decreased by unity and a constant is added and removed. The iterations stop as soon as the change of the model specification does not permit to diminish the corrected Akaike criterion. The results of the estimation are presented in Figure 3.4. For comparison purposes, we took hyperparameter set A for BVAR models as in Figure 3.1.

The Figure 3.4 shows that for some variables and some forecasting horizons the univariate models indeed outperform both random walk and multivariate VAR and BVAR models. However, it does not change the general conclusion that BVAR perform better than other models in terms of the forecast accuracy.

3.7 Conclusion

This paper evaluates the forecasting performance of BVARs on Russian data. We estimate BVARs of different sizes and compare the accuracy of their out-of-sample forecasts with those obtained with unrestricted VARs, ARIMA, ETS, and random walk. Our sample consists of 23 variables and we forecast at 5 different horizons up to 12 months. We show that for the majority of the variables BVARs outperform the competing models in terms of forecasting accuracy. However, we cannot confirm the conclusion drawn in some other studies (for example, Bloor and Matheson (2010), Bańbura et al. (2010)), where Bayesian methods were applied to data from developed countries, claiming that

	h=1	h=3	h=6	h=9	h=12
ind product	0,92		0,96	0,82	0,7
cpi	0,19	0,38	0,46	0,38	0,33
interb rate	0,96		0,66	0,52	0,58
agriculture	0,93	0,82	0,7	0,67	0,57
construction	0,97	0,95	0,99		0.98
employment	0,67	0,42	0,43	0,6	0,72
export	0,59	0,61	0,76	0,81	0,89
gas price	0,73	0,43	0,22	0,29	0,5
gov balance	0,6	0,79	0,77	0,7	0,63
import	0,75	0,48	0,52	0,72	0,98
labor request	0,66	0,79	0,94	0,95	0,96
lend rate	0,94	0,84	0,77	0,77	0,8
M2	0,46	0,51	0,63	0,72	0,83
nominal ER					
NFA of CB	0,6	0,56	0,75	0,65	0,6
oil price	0,88	0,81	0,88	0,81	0,77
ppi	0,43	0,75	0,69	0,59	0,49
real income	0,87	0,84	0,83	0,71	0,73
real invest	0,78	0,61	0,73	0,88	0,91
real ER	0,72	0,68	0,6		
retail	0,62	0,39	0,4	0,59	0,71
unemp rate	0,93	0,83	0,9	0,92	0,94
wage	0,74	0,51	0,46	0,42	0,41

Random walk
VAR3/4 - BVAR 3/4
VAR6/7
BVAR6/7
BVAR23
ARIMA
ETS

Figure 3.4: RMSFE of the best forecasting accuracy models, parameter set A: σ_i are std of $AR(p)$ residuals, $\delta_i = 1$ for nonstationary series, $\delta_i = 0.5$ for stationary series

high-dimensional BVARs forecast better than low-dimensional models. Our results imply that a 23-variable BVAR performs most accurately in only about a half of the cases where a BVAR is considered as a better forecasting tool with respect to its competitors. For the rest of those cases a BVAR with a relatively small size (6 or 7 variables in our case) can outperform a 23-variable BVAR in terms of forecasting accuracy.

Our robustness check results are mixed. On one hand, the BVAR specifications are included in the superior set more often than the VAR specifications. On the other hand, for most of the variables and forecasting horizons the forecasting performance of the RW model is not significantly inferior to that of BVARs according to the MCS procedure.

Appendices

.1 Data description

Name of serie	Type of series	Base period (if any)	Source
Industrial production index	base index	2010	IFS
Consumer price index	base index	2010	IFS
Employment in manufacturing index	base index	2010	IFS
Interbank interest rate	perc. per ann.		IFS
Lending interest rate	perc. per ann.		IFS
Real income index	base index	01:1992	FSSS
Unemployment rate	percent		IFS
Crude oil (Brent) price index	base index	2010	IFS
Producer price index	chain index		IFS
New houses commissioning	thous. of sq. met.		FSSS
Real fixed investment index	base index	01:1994	UAESD
Real wage rates index	base index	01:1993	FSSS
Monetary aggregate M2	bln. rub.		CBR
Real effective exchange rate	base index	2010	IFS
Natural gas price	US$ for bln BTU	2010	IFS
International reserves excluding gold	Bln US$		IFS
Nominal exchange rate	rub. per US$.		IFS
Declared need in workers	thous. of people		UAESD
Real agricultural production index	base index	01:1993	UAESD
Real retail output index	base index	01:1994	UAESD
Total government budgetary balance	bln. rub.		UAESD
Export of goods	mln US$		IFS
Import of goods	mln US$		IFS

IFS - International Financial Statistics of IMF http://www.imf.org/en/Data

FSSS - Federal State Statistical Servicehttp://www.gks.ru/

CBR - Central Bank of Russia http://cbr.ru/

UAESD - United Archive of Economic and Sociological Data http://sophist.hse.ru/rstat/

Chapter 4

DSGE-based forecasting: what should our perspective be?[1]

4.1 Introduction

In modern macroeconomics, modelling an economy in a dynamic stochastic general equilibrium (DSGE) framework attracts special attention. Significant progress in the specification and estimation of the DSGE models over the past 20 years has led to significant interest in them from the academic community and from central banks. The DSGE models can be used to determine the sources of business cycles, to forecast macroeconomic indicators, to analyse the effects of different policies and structural changes in the economy, etc. Besides, DSGE models are microfounded and consequently not susceptible to Lucas critique. On the other hand, being a stylized description of reality, the DSGE models obviously cannot reflect all the existing relationships between macroeconomic variables. But is this weakness essential for their forecasting performance? This paper compares the accuracy of forecasts made using a DSGE model and vector autoregressions (frequentest and Bayesian) on Russian quarterly data. On the basis of mean-square forecast errors (MSFE) I conclude that the DSGE model is usually inferior to BVAR model in terms of forecasting accuracy but the difference is not too large. At the same time, DSGE model allows the user to obtain a forecast with the minimal MSFE for some variables and some forecast horizons considered. Almost just after they were introduced in the

[1] published as Malakhovskaya (2016)

early 1980s, vector autoregressions became one of the main tools for forecasting and structural analysis and remain so until now, while the advantages of empirical analysis of transmission mechanisms and forecasting using DSGE models became apparent only relatively recently. In particular, it has been shown by Christiano et al. (2005) that an optimization model with nominal and real rigidity can well mimic the effects of monetary shock. Smets and Wouters (2007) show that the new Keynesian DSGE models can fit the dynamics of macroeconomic variables and forecast not worse (and even better in some cases) than BVARs. It is worth noting that New Area-Wide Model (Christoffel, Coenen, and Warne (2008)), a model that is currently actively used by the European Central Bank for forecasting and economic policy analysis is based on the papers by Smets and Wouters (Smets and Wouters (2003), Smets and Wouters (2007)). However, a good forecasting performance of DSGE models is not confirmed in all samples. For example, Edge and Gürkaynak (2011), conclude that the forecasting accuracy of several tools, including BVAR and DSGE, is low. Currently, there are several different published DSGE models designed to fit the dynamics of Russian macroeconomic indicators and estimated on Russian data (Malakhovskaya and Minabutdinov (2014), Polbin (2014), Shulgin (2014), Ivashchenko (2013)). Each model has its own advantages, but none of the papers, except Ivashchenko (2013), compare the quality of out-of-sample forecasts with non-structural models, which could be a good criterion of the successfulness of the model in question. Ivashchenko (2013) reports that the out-of-sample forecast based on DSGE model is more accurate that one based on competing models (frequentest AR and VAR models) for almost all the variables he considers. However, the author does not compare the quality of the DSGE model with a BVAR model that is a modern leading tool for macroeconomic forecasting. Moreover, this paper does not present the forecast errors separately for each forecasting horizon. A working paper of the Bank of Russia (Kreptsev D.A. (2016)) compares the quality of DSGE and BVAR forecasts, and analyses different specifications of their DSGE-model for two exchange rate regimes. The authors conclude that their DSGE with fixed exchange rate provides a more accurate forecast for three out of four variables considered, while the forecasting performance on the model with floating exchange rate is approximately the same as that of BVAR. This chapter uses the DSGE model presented in Malakhovskaya and Minabutdinov (2014) (chapter 1). For the reader's convenience, the interrelationships between sectors and markets in this model are described here, while a more detailed description of optimization problems can be found in the original paper. The model is an extended version of the models proposed by Kollmann (2001) and Dam and Linaa (2005)). A key feature of the extended version is the explicit modelling of revenues from energy exports, which reflects the export orientation of the Russian economy. Indeed, Russia is one of the largest world oil producers. Consequently, changes in the price of energy resources in the world market may be an important source of business fluctuations, and the oil market time series may contain significant information for forecasts of the Russian macroeconomic indicators. The Bayesian VAR model, which is used in the paper as a competing model for the DSGE in terms of forecasting accuracy, is estimated as by (Smets and Wouters (2007)).

The next section presents the main blocs of the DSGE model. Section 3 describes the estimation techniques. Section 4 presents the competing non-structural models. Section 5 reports the results of out-of-sample forecasting using DSGE, VAR and BVAR models. The Section 6 concludes and sketches some possible directions for future research.

4.2 DSGE model description

This section presents a theoretical model based on the paper by (Kollmann (2001)), some differences from Kollmann's model follow (Dam and Linaa (2005)) that in general also follow Kollmann (2001).

The benchmark model describes the connections among four economic agents: households, firms, the central bank and the foreign sector. As written above, an important feature of the model analyzed in the paper is explicit modeling of oil flows. Revenues from oil exports are assumed to be exogenous and go directly to the households as windfalls as in the paper by (Batte, Bénassy-Quéré, Carton, and Dufrénot (2009)).

4.2.1 Households

The population in the model is assumed to consist of many households. A representative household maximizes the expected present value of utility over an infinite horizon with an infinite set of budget constraints. Preferences are defined on the consumption and

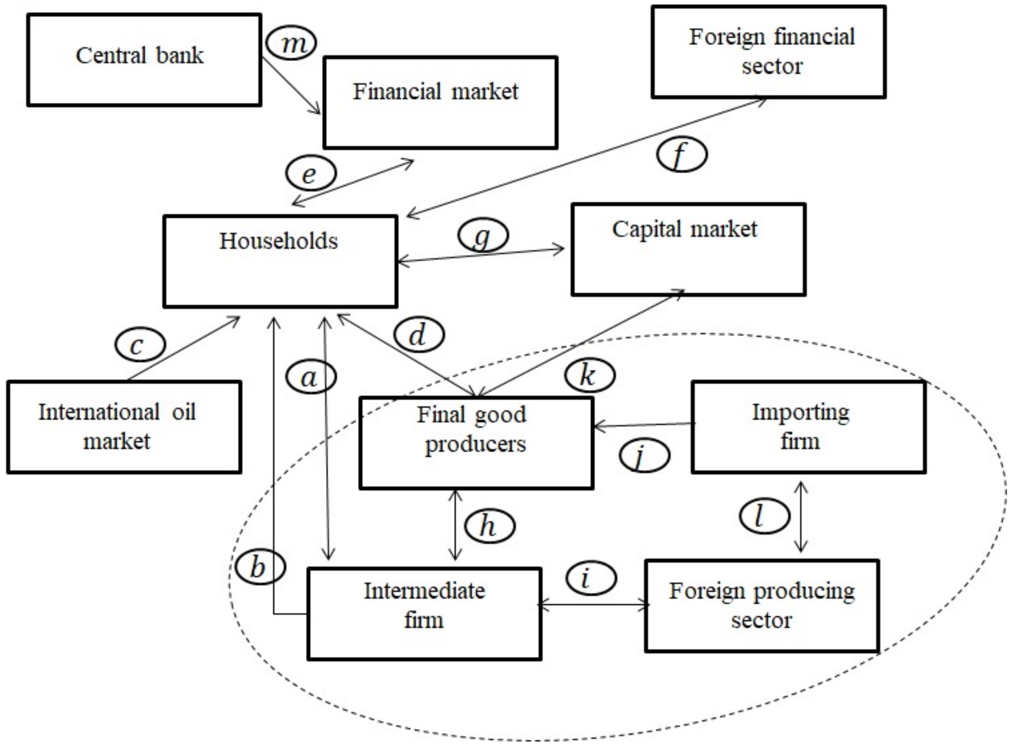


Figure 4.1: The scheme of the economy

labour services space. Following many papers, for example Erceg et al. (2000) Gali (2008), it is assumed that labour services are differentiated and each household is a imperfectly competitive supplier in the labour market and provides labour services to each of the firms operating in the intermediate good market (arrow a on Figure 1). Households can set the wages. It is assumed that household wage setting, as well as firm pricing, goes following the Calvo mechanism (Calvo (1983)) with some probability of receiving a wage change signal. A household meets the demand for its labour at a fixed wage, regardless of whether it has the right to change the wage in the period or not.

Households can have three types of assets which are domestic and foreign bonds and capital stock. A household receives an income from its assets, labour income [2] (arrow a) and dividends from imperfectly competitive firms producing intermediate products (arrow b), as well as income from energy exports (arrow c). The household distributes the income between consumption (arrow d), new investments in financial assets (arrow e,f) and capital

[2]A bidirectional arrow means that there are financial and/or real flows from one agent to another and vice versa. In this case, for example, a bidirectional arrow shows that labour services are provided by households to firms and the households receive wages in return

(arrow g). The imperfections in the financial market are assumed to create deviations in the interest rate on foreign bonds (the negative value of foreign bonds is the debt of households to the foreign sector), which can be interpreted as a risk premium. Thus, if the debts of households due to the foreign sector increase, the interest rate for households increases as well. Technically, an endogenous risk premium garantees a stationary equilibrium in the model (Schmitt-Grohe and Uribe (2003)). Each period of time the household solves the problem of intratemporal choice between work and leisure, while the household chooses a new wage only if it receives an information signal.

4.2.2 Productive sector

Modelling the manufacturing sector generally follows the papers by (Kollmann (2001) and Curdia and Finocchiaro (2005). Intermediate goods in the model are traded, while the only final good produced by aggregating intermediate domestic and imported goods is not traded. Intermediate goods of different firms are assumed to be imperfect substitutes for each other. In Fig. 1, the production sector is encircled with a dotted line for the reader's convenience.

Intermediate goods production

An intermediate good of a representative firm is produced with capital and labour services. The intermediate firm uses differentiated labour services from all households, while the capital is assumed to be homogeneous. Accordingly, wages may be different for different households. Relatively higher wages of a certain household in comparison with other households lead to lower demand for its labor services from intermediate firms. It is assumed that goods of intermediate firms can be sold in the domestic market (arrow h) or exported (arrow i). Intermediate goods sold in the domestic market are supplied to firms producing final goods. Intermediate producers are able to discriminate in terms of price, i.e. the price of an intermediate good of a firm within a country does not necessarily equal the price of the exported good of the same firm, adjusted for the exchange rate. The assumption of possible price discrimination (in this case discrimination is the result of behaviour described as pricing-to-market (Knetter (1993))[3] is based on the extensive

[3] And as a consequence, there is a violation of the law of one price and purchasing power parity.

literature, both theoretical and empirical (see, for example, (Balassa (1964), Taylor and Taylor (2004)), showing that the absolute PPP does not hold, at least in the short run. This model assumes that the prices of both imported and exported goods are sticky, and that they are set in the currency of a purchaser. In other words, not only intermediate producers (and exporters) are imperfectly competitive and can set the price of their products, but also importers are. As in the case of households, pricing is 'a la Calvo and a firm can only change its price in a period with a certain probability. By setting a new price, the firm maximizes the expected discounted profit on an infinite horizon.

Final good production

Unlike the market for intermediate products, which is assumed to be imperfectly competitive, the market for final products is assumed to be perfectly competitive. Technically speaking, the imperfect competition in one of the two goods markets (final or intermediate) is enough to have the non-neutrality of monetary policy in the model in the short run. Since the sector of final goods is absolutely competitive, the number of firms in this sector does not make a difference. One could even assume that there is just one firm in this market which functions as a perfect competitor and receives zero profit. An absolutely competitive firm produces the final good using the intermediate domestic (arrow h) and imported goods (arrow j) without additional costs. The final good can be used for household consumption (arrow d) or as capital (arrow k).

4.2.3 Foreign sector

Export and import

The model assumes that exports are carried out by domestic intermediate firms, and the goods are used in the foreign productive sector (arrow i), with the sane economic structure as in the domestic economy. Intermediate foreign products are imported from abroad by foreign firms, with each firm importing its own type of product. Imported goods are produced by the foreign manufacturing sector (arrow l). As well as domestic intermediate goods, any imported product is assumed to be an imperfect substitute for others. Imported intermediate products are absorbed by the final goods sector (arrow j) and used in the final good production process. As in the domestic market, in order to be able to change the price of its products, an exporting or importing firm must receive a random price signal, which comes with a certain probability. When choosing a new price, the firm maximizes the expected discounted profit on an infinite horizon.

4.2.4 Central Bank

In the model, the central bank implements the monetary policy, setting the nominal interest rate. The yield of domestic bonds is equal to the nominal interest rate set by the central bank. The central bank is assumed to follow a Taylor rule, taking into account the current inflation and output. Different money policy rules embedded in the DSGE model may influence its forecasting performance. Since the purpose of the model is forecasting, it is preferable to use the rule reflecting the actual strategy of the central bank. At the moment of the writing (2016), the Bank of Russia is implementing an inflation targeting policy, which implies a particular set of rules suitable for describing its current strategy. However, the model is estimated on the data for more than 10 years, and there is no consensus about a relevant rule of the Bank of Russia for this time period in the academic literature.

In different volumes of the Bulletin of the Bank of Russia issued over the years one can read that the main task of the Central Bank is to reduce inflation (the exact targets vary slightly from year to year), while ensuring the stability of the national currency. In addition, every year the Bank of Russia expressed its intention to implement the inflation targeting [4]. In particular, in 2011 the Bank of Russia announced its intention to "use the short-term interest rate of the interbank credit market as the operational benchmark of the interest rate policy" (Bulletin of the Bank of Russia, 2011, p. 4). At that time, it is argued that the "value of the bi-currency basket" remained the operational target (Ibid., p. 8). Besides, the Central Bank of Russia published the expected money supply growth, computing it depending on the inflation target, however, assuming that in the future "the money supply growth will be due to the increase of the net domestic assets [...]

[4] At the end of 2012, it was announced that Russia would make the final transition to inflation targeting in 2015, but the intention to make this transition in the near future can already be found in the Bulletin of the Bank of Russia issued in 2008.

while the net international reserves will remain stable" (Ibid., p. 21). Thus, the Central Bank of Russia simultaneously pursued several goals and it is not clear which of them was the highest priority at any particular moment. At the same time, the central bank had a wide range of tools to attain its goals. So, it seems difficult to formulate a single policy rule that may well describe the monetary policy in Russia in the 2000s. The economic literature reflects the lack of consensus about the stance of monetary policy in Russia. For example, Vdovichenko and Voronina (2006) suggest that the monetary policy in 1999-2003 is better modelled with a money supply rule. On the contrary, Benedictow et al. (2013) estimate a model of the Russian economy using the data for 1995-2008 and conclude that the monetary policy follows a simple Taylor rule with an interest rate responding to changes in unemployment and inflation. The authors make a caveat that the assumptions underlying this rule are hardly applicable to the Russian economy but the model with a Taylor rule fits data well. Similarly, Taro (2010) assesses a non-linear interest rate rule for Russia using data from January 1997 to January 2007 making an assumption that the central bank preferences with regard to the output gap and inflation is asymmetrical. Finally, (Yudaeva et al. (2010)) show that a forward-looking interest rate rule, similar to the Taylor rule, may fit well the Russian data. They also replace the interest rule by a monetary aggregate in a monetary rule addressing an idea that the effect of the interest rate change on the economy is insignificant due to the weak development of financial markets. They do not draw a conclusion in favor of any of these rules. However, evidently there are papers showing that the monetary policy stance in Russia can be represented with an interest rate Taylor rule. Following these papers I model the monetary policy in the same vein. In the model, the central bank directly manages the return on domestic financial assets, its activity is indicated by arrow m in Fig. 1.

4.3 DSGE model estimation

The DSGE model outlined in the previous section is estimated using Bayesian methods. Bayesian estimation is a combination of maximum likelihood (determined by the model structure and the data) with a prior distribution to obtain a posterior distribution of the parameters of interest. The use of some prior information can raise questions about the sources of that information and their reliability. However, from a practical point of view, the use of a prior distribution generally improves the estimates. Moreover, the combination of a likelihood with a prior distributions is particularly perspective when a model is estimated on data from an emerging market economiy where the avalaible data series are usually short. When the sample size is small, the likelihood function can be almost flat, and combining it with a reasonable good distribution can help identify the model (Fernandez-Villaverde (2010). To estimate the model in the paper I try to choose the prior distribution that is not too narrow to avoid obtaining the posterior distribution just with the prior knowledge. The model is estimated using nine time series. The observed variables are consumption, domestic inflation, energy export revenues, domestic interest rates, real exchange rates, real wages, foreign interest rates, foreign output and foreign inflation. Most of the data is taken from the International Monetary Fund database. All series are quarterly, rolling from Q1 1999 to Q1 2016. I admit that the series are rather short, but earlier data are not used because of the 1998 crisis. At the same time, it is known that since 2008 the central bank has changed its policy, but the sample is not limited by the end of 2008 in order to avoid estimating the model with an even shorter series. The GDP deflator is used for deflation. The series that demonstrate seasonality are seasonally adjusted. Estimation and forecasting for all models is performed using only data that could have been available in case of real-time forecasting. For example, data up to the second quarter of 2013 are used to build a one-quarter forecast for the third quarter of 2013. The DSGE model is a stationary log-linearized model and requires preliminary data detrending [5]. Although the VAR and BVAR models can be estimated on the raw initial data, they are estimated using detrended data for comparability of results. The trend is also calculated recursively. Nominal household consumption expenditure is used as a series of consumption. After deflation, I seasonally adjust the series and calculate the per capita consumption [6]. The quarterly growth rate of producer prices is taken as a series of domestic inflation. The money market rate (interbank rate) is taken as a

[5]The pre-filtering procedure before the DSGE assessment is standard practice (see, for example, Smets and Wouters (2003), Lubik and Schorfheide (2007), Dib (2011), Justiniano and Preston (2010)). The alternative is to include the trend in the model.

[6]The population was taken from the Russian statistical database (Rosstat). The annual data is interpolated to a quarterly series using a spline function.

domestic interest rate. The source of data for real wages is the Rosstat database. The real value of the bi-currency basket against the rouble (weights are 0.55 for the US dollar and 0.45 for euro) is taken as an observable equivalent of the real exchange rate. This basket was used by the Bank of Russia as an operational target of the exchange rate policy from February 2007 to November 2014. [7] The real exchange rate of the US dollar and euro is calculated using consumer price indices. As a variable of oil prices, I take the revenues from the export of crude oil, petroleum products and natural gas. This is due to the fact that the model assumes a constant oil output and all changes in household income are due to changes in its price. Besides, the model assumes only one source of raw material exports. The use of all commodity export revenues instead of the price of oil also permits to take the natural gas into account. The data source about the export revenues is the balance of payments statistics provided by the Bank of Russia. The series are recalculated to be expressed in artificial currency, with weights of dollar and euro be equal to 0.55 and 0.45, respectively. Finally, per capita export revenues are calculated. All variables of the external sector are calculated as weighted average of the corresponding variables for the USA and Eurozone with weights equal to 0.55 and 0.45, respectively. The foreign inflation is calculated using consumer price indices. The foreign interest rate is calculated using money market rates. The foreign output is calculated using data on real GDPs. I take logarithms of the series of consumption, real wages, real exchange rate, revenues from oil exports and foreign output, and then the series are linearly de-trended. The interest rates were divided by 400 to make them quarterly. Finally, the series of interest rates and inflations are re-centered, giving them a zero mean [8]. Prior parameter

[7] The use of the bi-currency basket as an operational target started in February 2005, prior to which the US dollar had played that role. During two years, the shares of the dollar and the euro have been revised five times, each time to increase the share of the euro. The change of shares in the model is not taken into account intentionally in order not to create artificial spikes in the value of foreign sector variables. Moreover, the DSGE models are often estimated using an effective exchange rate. According to preliminary estimates, the average share of exports and imports from the Eurozone and Switzerland among Russia's 15 major trade partners was 45.2% between 1999 and 2011. (author's calculations according to the IMF data). So if an effective exchange rate were used for the estimation the weights of the euro and the US dollar would have been fixed at the same levels.

[8] The trend definition for the main variables reflects the neoclassical growth theory, according to which the variables such as consumption and wages per employee should have a constant growth rate,

distributions are chosen in a standard way. Several parameters are fixed (i.e. the prior distributions with a zero variance are chosen) following other similar papers. Most of the model parameters, including preference parameters, parameters of the production function, capital adjustment cost parameter, as well as autocorrelation coefficients and standard errors of structural shocks, are estimated. When choosing a prior distribution, a general rule applies: all parameters that can take values between zero and one have beta prior distribution, all preference parameters have prior gamma prior distribution, parameters of monetary rule have normal prior distribution and all standard errors of structural shocks have prior inverse gamma distribution. The convergence of parameters is achieved at each sample size, which is confirmed by the results of the special tests.

4.4 VAR and BVAR models

The competing models are reduced-form VAR and BVAR models estimated on the same data. The section discusses their specification.

Let the variables be combined into a m-dimensional vector. A reduced-form vector autoregression (VAR) model is written as:

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \varepsilon_t \qquad \varepsilon_t \sim N(0, \Sigma) \tag{4.1}$$

where Φ_0 is a vector of constants of dimension m, Φ_l are autoregressive lag matrices of dimension $m \times m$. Vector ε_t is a vector of errors of dimension m uncorrelated with explanatory variables.

A frequentist VAR can be estimated by OLS applied consequently to each equation. The Bayesian VAR means that the model is estimated using Bayesian methods. A frequentist VAR is a special case of Bayesian, for simplification I will refer them in what follows as two different models. To be able to compare the forecasts from three models in hand, VAR and BVAR are estimated using the same 9 variables, which serve as observables for the DSGE estimation.

while inflation and interest rates should remain constant on the balanced growth path. From a technical point of view, the use of a deterministic trend avoids unreliable assumptions about the trend dynamics out-of-sample

The prior distribution for the baseline BVAR model is chosen following Smets and Wouters (2007), which also follows Sims and Zha (1998) and Sims (2003). The prior distribution is defined using three blocks of dummy observations (for an overview of the BVAR method and a method of defining a conjugate normal - inverse Wishart (cNIW) prior distribution with dummy observations see chapter 2). Technically, I combine three different prior distributions: cNIW prior, sum of coefficients prior and initial observation prior. Such a combination of prior distributions have shown a good forecasting performance on different samples (for the United States, see, for example, Robertson and Tallman (1999), Bańbura et al. (2010), Carriero et al. (2015), for Russia, see chapter 4). Hyperparameters of the prior distribution were chosen following Smets and Wouters (2007), but contrary to that paper I did not perform a prior distribution correction with a training sample due to the short length of available time series.

4.5 Forecasting

Forecasting with all three models is done on an expanding sample. The shortest training sample consists of 50 observations and ends in the second quarter of 2011. Correspondingly, this period is the earliest one used for forecasting. The second training sample consists of 51 quarters, and the forecast is made in the third quarter of 2011, etc. The out-of-sample forecasting horizon is from one to eight quarters. This method of estimation and forecasting allows to obtain 19 estimation samples and to make 19 one-period-ahead forecasts, 18 two-period-ahead forecasts, etc. For each model and for each variable the mean square error of the forecast (MSFE) is calculated according to the formula:

$$MSFE^{M}_{var,h} = \frac{1}{N_h} \sum_{T} (y^{M}_{var,T+h|T} - y_{var,T+h})^2, \tag{4.2}$$

where $y^{M}_{var,T+h|T}$ is a forecast for the variable var, made with a model M at the moment T for h steps ahead. $y_{var,T+h}$ is an actual value of the variable for the moment $T+h$. A number of the forecasts made at different T for the forecasting horizon h is denoted by N_h. In this case, T goes from 50 to 68, h goes from 1 to 8, var is one of the nine observed variables described in the previous section, M is one of the three models in hand $M = \{VAR, BVAR, DSGE\}$ and N_h can take values from 12 to 19, depending on the forecasting horizon. Obviously, the smaller the MSFE, the more accurate the forecast is. MSFE values for all three models considered and all forecasting horizons are shown in Table 1 [9]. To save space, only variables directly related to the Russian economy are presented. The variables of the foreign sector are weighted averages of the corresponding variables of the Eurozone and the USA, and their forecasting is of little interest for this chapter. In the table below, the values corresponding to the model with the best forecasting performance among three models in hand are given in bold. In a baseline model the order of VAR and BVAR is equal to two that follows Smets and Wouters (2007). To check the robustness of the results, the number of lags for the VAR and BVAR models is also selected by maximizing the data density function (marginal likelihood). Its maximum value is achieved at 1 lag in the case of VAR (for 18 data sets from 19) and 7 lags in the case of BVAR[10] (for 16 data sets from 19). VAR(1) and BVAR(7) models permit to forecast some variables on some horizons more accurately than VAR(2) and BVAR(2), respectively. As a result, the DSGE model has given the most accurate forecast of the domestic inflation on a 12-month horizon but it did not change the model ranking on short horizons, so to save space, the table presents the results for the baseline case only. Table 1 shows that for most of the variables and forecast horizons, the model with the smallest forecast error is BVAR, which confirms the practice of making forecasts using this model in the academic literature.

For some variables and some forecasting horizons, the DSGE model is more accurate than the BVAR model. At the same time, some interesting features can be identified. For example, it is preferable to do a very short-term forecast of the interest rate using the BVAR model while for longer horizons the DSGE model performs better. An opposite "rule-of-thumb" applies to the domestic inflation. The frequentist VAR does not give the most accurate forecasts for any variable and any forecasting horizon, which partly explains a current tendency of abandoning the frequentist VAR models in favor of Bayesian models

[9] Table 1 shows the MSFE for the detrended data. Since the detrending here is a monotonous transformation, the ranking of the models stays the same if MSFE is calculated on the initial data with trend. The DSGE forecast is an average point forecast that takes into account the uncertainty of both shocks and model parameters.

[10] Marginal likelihood values were compared for models with the number of lags up to four for VAR and up to ten for BVAR.

Variable	Model	1q.	2q.	3q	4q	5q	6q	7q	8q
real	DSGE	0.118	0.187	0.248	0.298	0.355	0.400	0.435	0.467
exchange	VAR	0.123	0.186	0.253	0.343	0.312	0.378	0.432	0.473
rate	BVAR	0.102	0.142	0.182	0.212	0.271	0.338	0.379	0.420
internal	DSGE	0.024	0.035	0.038	0.041	0.040	0.040	0.039	0.037
inflation	VAR	0.032	0.039	0.043	0.049	0.047	0.035	0.034	0.033
	BVAR	0.026	0.039	0.046	0.057	0.036	0.028	0.021	0.025
interest	DSGE	0.006	0.009	0.010	0.011	0.011	0.011	0.012	0.013
rate	VAR	0.008	0.012	0.014	0.017	0.015	0.016	0.015	0.015
	BVAR	0.005	0.008	0.011	0.013	0.014	0.016	0.017	0.018
oil	DSGE	0.195	0.322	0.450	0.554	0.649	0.722	0.792	0.855
revenues	VAR	0.138	0.265	0.436	0.553	0.653	0.816	0.924	0.996
	BVAR	0.136	0.218	0.352	0.441	0.597	0.708	0.801	0.879
real	DSGE	0.044	0.094	0.144	0.191	0.234	0.269	0.303	0.333
wages	VAR	0.031	0.064	0.104	0.162	0.160	0.202	0.244	0.274
	BVAR	0.025	0.043	0.065	0.095	0.122	0.152	0.178	0.201
consumer	DSGE	0.024	0.050	0.073	0.096	0.117	0.131	0.143	0.154
spendings	VAR	0.021	0.033	0.042	0.061	0.070	0.091	0.109	0.122
	BVAR	0.015	0.023	0.031	0.045	0.056	0.069	0.080	0.092

Table 4.1: RMSFE for DSGE, VAR and BVAR models

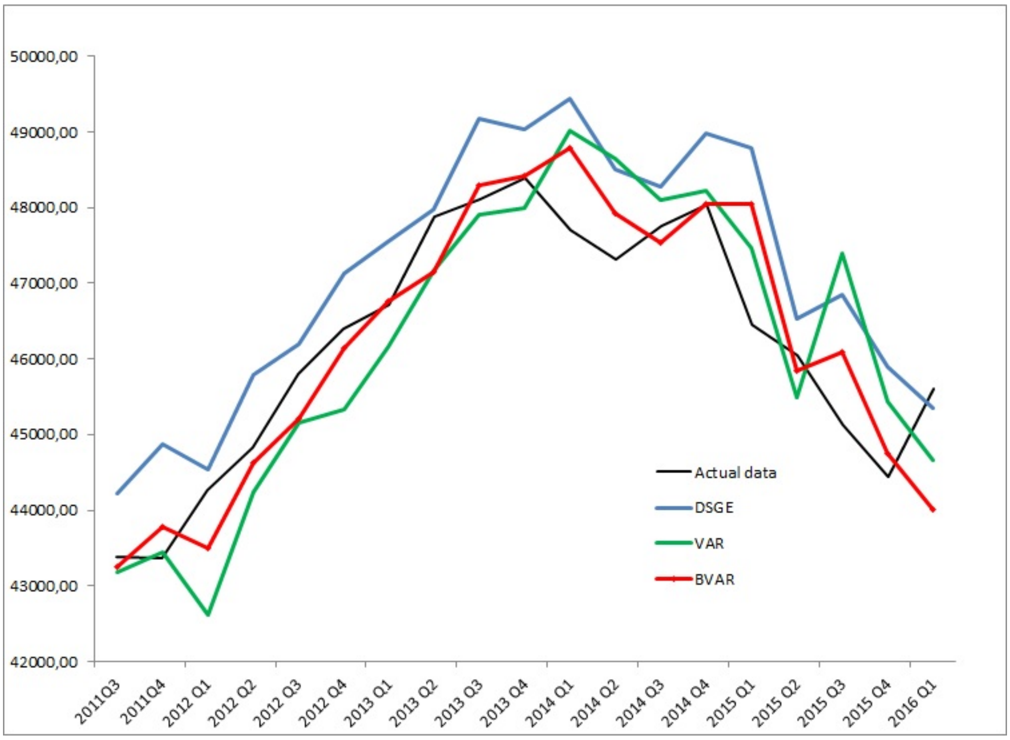


Figure 4.2: Actual consumer spendings and forecasts using three models (raw untransformed data)

for both forecasting and structural analysis.

Then I analyze the dynamics of forecasts for the consumer spendings variable for which the DSGE model gives the least accurate forecast for all forecasting horizons. Fig 2. shows the actual dynamics of this variable and its one-quarter ahead forecasts made with all three models under consideration. To obtain the series presented in Fig. 2, the forecasts were made on the detrended data, and then the trends were restored. The unit of measurement is real rouble, the base period is 2010. The forecasts for all three models capture the general trend well, but all three model fail to predict the short-term fluctuations (one to three quarters), and react to them with a delay. A peculiar feature of the forecast made with the DSGE model is that all forecast values are above of the actual values (which partially explains the low quality of forecasts on this model). This result may seem counter-intuitive but it is due to the fact that recursive detrending is used for estimation and forecasting. It is worth noting that the mean forecast error increases significantly with the forecasting horizon for almost all variables and models. For the eight-quarter ahead

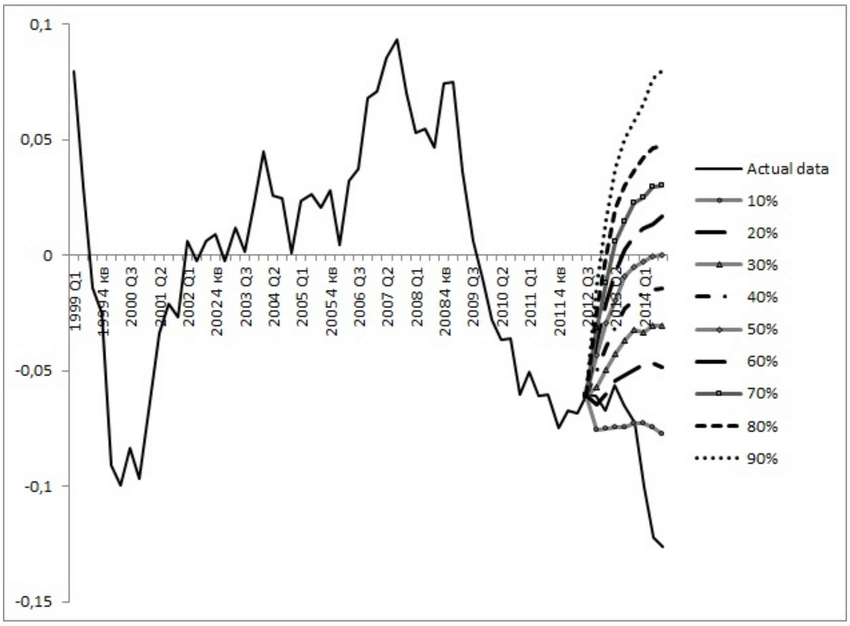


Figure 4.3: Actual consumer spendings and deciles of forecasts of this variable by DSGE model (data are presented in deviation from the log-linear trend).

forecast the MSFE is 4-6 times bigger than for one-quarter ahead forecast. It means that the accuracy of all forecasts for the horizon for more than one year is very low. Fig. 3 shows the dynamics of the detrended series of consumer spending (the trend is calculated on the sample from Q1 1999 to Q3 2012) and forecasts made with the DSGE model with data up to Q3 2012[11] for all horizons from one quarter ahead to eight quarters ahead. Alternative forecasts correspond to nine deciles of forecast density, i.e. 10% of the lowest forecasts are below the first decile line, 20% of the lowest forecasts are below the second decile line, etc. The distribution of forecasts shows that as a general rule the DSGE model predicts a return to the stationary state (that is equal to zero in this case). This is not surprising, as the DSGE model is a business cycle model and it implies implicitly the end of the recession and return to the natural level.

However, actually the recession did not end, and even deepened. It explains why the average and median forecasts regularly exceed the actual value. At the same time, it is impossible to make forecasts for the model on Russian data without a permanent recession in the testing sample. The available data begin in the Q1 1999, and it does not make sense

[11] The period is chosen randomly

Model	1q.	2q.	3q	4q	5q	6q	7q	8q
DSGE (second decile)	0.014	0.025	0.042	0.060	0.077	0.090	0.101	0.111
DSGE (mean forecast)	0.024	0.050	0.073	0.096	0.117	0.131	0.143	0.154
BVAR	0.015	0.023	0.031	0.045	0.056	0.069	0.080	0.092

Table 4.2: RMSFE for consumer spendings using DSGE and BVAR models

to estimate the model using less than 40 observations, so the earliest possible forecast starts in the Q1 2009, which is in the acute phase of the financial crisis. For examole, the Fig. 3. presents the third quarter of 2009 as a threshold between the periods of positive and negative deviations of consumer spending from the trend, but the same period shows a recession if the data is detrended with the trend calculated on a shorter sample. If the forecast is done using detrended data up to Q3 2009 the result is similar to one described above: the mean point forecast implies a gradual exit from the recession, while actually the recession deepened. It is worth noting that the consumer spending forecast corresponding to the second decile of the DSGE model gives a much lower MSFE compared to the mean forecast. The forecast performance of the DSGE model (for 1 or 2 quarters ahead) is approximately the same as the forecast performance of the BVAR model (see Table 2).

So, a relatively inaccurate forecast with the DSGE model can be hardly regarded as a methodological shortcoming of the DSGE model itself. It is rather a certain kind of a bad luck: most of the training sample contains a period of stable growth, whereas in 2009 the data show a flatter trend, and in 2014 consumer spending start to decrease in real terms in the sample in hand. At the same time, it is impossible to test the model's forecasting performance on a more homogeneous sample due to the lack of long time series of Russian data.

4.6 Conclusion

Microfounded DSGE models have become an important part of the modern macroeconomics. They are in an active use both by the academic community and central banks of developed countries and to serve to address a question of an impact of various shocks or policy measures on the economy, as well as to make forecasts. This paper analyzes a relative accuracy of the DSGE forecasts compared to ones made with the reduced form VARS (frequentist and Bayesian). The DSGE model used in the chapter accounts the revenues from commodity export explicitly. For estimation of the BVAR model I use a modification of the conjugate normal - inverse Wishart distribution which a regular practice in the literature. Based on the mean-square forecast errors, I conclude that the accuracy of forecasts made with the DSGE model is generally lower than that of forecasts made with the BVAR model. However, the DSGE model provides the most accurate forecast for some variables and some forecasting horizons among three models in hand. At the same time, there are data-related reasons why DSGE fails to accurately forecast some variables in the sample, such as consumer spendings. As a possible line for further research, it seems perspective to make a DSGE with an explicit modelling of a trend. It would make possible to do the forecasts on the raw initial data without a necessary preliminary transformation, which might improve the forecasting accuracy of the DSGE model.

Chapter 5

Oil market shocks effects on Russian macroeconomic indicators: quantitative estimates with sign-identified SBVAR

5.1 Introduction

The influence of world oil market events on the Russian economy is regularly discussed in media and in popular economic literature. The evident reason of the interest is the large share of Russian exports that is attributable to crude oil and oil products. An oil price drop decreases the trade and the budget balance. Therefore, it may be surprising that quantitative estimates of oil price effects on macroeconomic indicators in Russia are extremely scarce. This paper fills that gap and aims at quantifying the effects of different oil market shocks on three key macroeconomic indicators in Russia. To reach the goal, in this paper I construct and estimate a structural Bayesian VAR model (SBVAR) and identify it with sign restrictions. The results are based on impulse response functions (IRF) analysis and forward error variance decomposition.

The shocks identification in the paper is realized as proposed by Baumeister and Hamilton (2015b). Baumeister and Hamilton (2015c) show how the oil shocks identified following the same algorithm affect oil market variables. However, no previous papers have shown how the shocks identified in the same way affect external variables with respect to the oil market. As empirical research, this paper is of one the rare examples of econometric analysis of oil market shocks on the Russian economy, and the first one that uses the SBVAR model.

The impulse response functions analysis show mixed results for the real monetary incomes and CPI inflation but they demonstrate that two of three oil demand shocks analyzed in the chapter positively affect the industrial production index.

The mean estimate of the forward error variance explained by oil market shocks at one-year horizon is between 14 and 30% for real monetary incomes and between 15% and 25% for inflation depending on the prior distribution. The fraction of the forward error variance of industrial production explained by oil shocks at a one-year horizon is between 35% and 45%. These values seem to be surprisingly high even for an oil-dependent economy like Russia's and need to be examined using a sensitivity analysis.

5.2 Literature review

This paper continues two strands of macroeconomic literature. On one hand, the paper complements the existing research on the impact of oil market shocks (more broadly, the impact of changes in natural resources prices) on the macroeconomic performance of different countries. At the same time, the existing papers usually analyse the impact of shocks on the developed economies (primarily, the United States), which are mainly oil importers. There are very few studies devoted to the impact of oil shocks on exporting countries. On the other hand, the paper uses a new method of imposing sign restrictions proposed by Baumeister and Hamilton (2015b) and continues the discussion about the relevant identification restrictions.

5.2.1 Influence of oil shocks.

One of the first key works on the impact of oil prices on key macroeconomic indicators is the book by Bruno and Sachs (1985), which analyses in detail the impact of oil shocks on a group of developed countries. Hamilton (1983) shows that nine out of ten recessions in the U.S were preceded by oil price shocks and Hamilton (1996) also demonstrates a significant impact of oil prices on the U.S. economy. These studies confirm that the oil price shock can be regarded as a global one that can trigger a new phase of a business cycle in many countries in the same time. On the other hand, Hooker (1996) challenges this opinion and shows that the severity of oil shocks tends to decrease. In this paper, the author concludes that, unlike the oil shocks of the 1970s, the oil price dynamics of the 1980s did not have a significant impact on the oil sector and, accordingly, the oil price itself has a low effect on the key US macroeconomic indicators. For European countries, Mork et al. (1994) and Cuñado and Pérez de Gracia (2003) conclude that the real effects of oil price changes have been reduced since the mid-1980s. However, the assumption about complete exogeneity of the price of oil seems to be irrelevant, especially for large importing countries. Barsky and Kilian (2002) and Barsky and Kilian (2004) suggest a structural explanation of the correlations found in previous papers. They show that the fluctuations in oil prices are not exclusively related to the oil producing countries of the Middle East. On the contrary, their analysis shows that the decisions of OPEC countries are a reaction to the global macroeconomic processes that determine the demand for oil. Thus, these papers explain the need for a separate analysis of the impact of supply of oil and demand of oil shocks. In case of separate analysis of those shocks, the data clearly show that the role of supply shocks has diminished over time (Edelstein and Kilian (2007), Herrera and Pesavento (2009)).

Using a structural model, Blanchard and Gali (2008) confirm the conclusion of Hooker (1996) and extend it to several developed countries (France, Great Britain, Italy, USA). The authors show that the impact of oil on output and inflation declined over time in these countries. However, the data for Germany and Japan show a different pattern. The reasons for the apparent reduction in the importance of oil shock as a source of business fluctuations are discussed in Baumeister and Peersman (2013). They note that since the 1980s there has been a reduction in the elasticity of demand in the oil market. Thus, the same price increase caused by the negative supply shock causes a smaller drop in world production and therefore has less impact on macroindicators. However, the impact of changes in global oil production on GDP growth remained the same. According to their estimates, the ratio of the forward error variance of inflation explained by the oil supply shock varies from 15 to 20%. Kilian (2009) is the first to model the oil market itself in place of using some general oil price shock. The paper identifies three types of shocks potentially affecting the oil price and production (global demand shock, demand shock in the oil market (demand for precautionary reasons) and a supply shock. The paper shows that price changes have different effects on the U.S. economy depending on the type of shock that causes the change. In particular, a price increase caused by a positive global demand shock has less negative consequences than a negative supply shock. The publication of Kilian (2009) resulted in a surge of research in this area and, thus a large number of papers analysing different aspects of the oil market appeared. Most of these papers like Kilian (2009), identify several oil market shocks separately. This separation certainly makes sense as the estimated impact of the different shocks happens to be different (Lippi and Nobili (2012),Aastveit (2014)). For example, Stock and Watson (2016) show that a supply shock explains a small proportion of output forecast errors, while demand shock explains a large proportion of oil price forecast errors. As this review shows, the impact of the oil shock on the U.S. economy and other developed countries, mainly oil importers, is frequently studied, while the impact of oil market shocks on emerging and oil-exporting countries is much rarely analysed in academic literature.

The impact of oil price shocks on an emerging market economy (an oil importer, though) is studied by Cavalcanti and Jalles (2013). Their results show that the impact of oil shocks on GDP growth in Brazil is not significant, and the oil shocks explain only a small part of inflation volatility. The authors explain the results by a decrease in the share of oil imports from 80% in 1980 to less than 5% in 2007 as a result of a program of replacement gasoline with ethanol produced from sugar cane.

The idea of asymmetric impact of oil market shocks on exporting and importing countries is studied by Mork et al. (1994). Their sample consists of six countries including two net exporters for most of the time (Norway and the UK). The authors study the asymmetric impact of positive and negative oil price shocks. Of the 6 countries in their sample, only Norway and the UK were net exporters for most of the sample. For most of the countries, the data show the negative impact of the oil price shock on economic activity, with one exception, and that is Norway.

At the same time, the strength of the effect of the shock varies from country to country: in the USA the effect of the oil shocks was stronger than in France, Germany and Japan, although the USA were less dependent on imported oil during the estimation period. At the same time, the UK responded to the oil shocks as a net importer, not as a net exporter. In Norway an oil shock has a positive impact on the output. The possible cause is a high share of oil exports.

A rare example of a study of the world naturual resources price shocks effect on an exporting economy is presented by Charnavoki and Dolado (2014). The authors use the Canadian economy as an example. The authors show that in a large extent the resources price fluctuations are determined by global demand shocks and specific resource supply shocks. The authors conclude that a positive resources price shock has a positive effect on the current account regardless of the type of shock. The Dutch disease can only be caused by a negative supply shock.

5.2.2 VAR identification schemes

The first structural vector autoregressive models used for oil shock analysis (for example, Kilian (2009) and Blanchard and Gali (2008)) imply recursive identification. As it is well known, the recursive scheme has a number of serious drawbacks (see Amir Ahmadi and Uhlig (2015) for a detailed overview). In particular, the model is obviously not properly specified if the variable assumed to be «most exogenous» in reality reacts to changes in other variables included in the model within one period.

For example, a recursive scheme may misrepresent the structure of the economy if the price of oil is not exogenous to other variables. The endogeneity of oil prices has been confirmed in a large number of studies. For example, He, Wang, and Lai (2010) show that oil futures prices are cointegrated with the U.S. economic activity index and the U.S. trade index, with the Granger causality from the economic activity to the oil price.

The papers that introduce identifying supply shocks in the oil market with sign restrictions are Peersman and Van Robays (2009) and Baumeister et al. (2010), where the authors analyse the impact of shocks on the economic performance of different countries using a partly identified model.

Their implicit criticism of the recursive scheme fostered Kilian and Murphy (2012) to re-estimate the model proposed by Kilian (2009), but with sign restrictions, the conclusions of this work remain unchanged. After the papers by Peersman and Van Robays (2009) and Baumeister et al. (2010) using of sign restrictions for oil market shock analysis has become very popular. For example, Baumeister and Peersman (2013) combine sign identification with time varying parameters, as in Primiceri (2005). Similarly, sign restrictions are used by, for example, Melolinna (2012), but in that paper the restrictions are applied in a non-standard way and using a penalty function.

The most serious criticism of traditional sign restrictions, is laid by Baumeister and Hamilton (2015b). They show that in the case of usual sign restrictions, a posterior distribution of simultaneous coefficients is determined by a prior distribution and the covariance matrix of residuals. This means that the impulse response function within the identifying set of parameters are determined by an implicitly specified prior distribution and does not depend on the likelihood function. The authors introduce an explicit prior distributions for coefficients with economic interpretation (e.g., elasticity of supply and demand). They apply their algorithm to identify labour market shocks (Baumeister and Hamilton (2015b)), monetary policy (Baumeister and Hamilton (2015a)) and the oil market (Baumeister and Hamilton (2015c)). The model pesented in this chapter is an extension of the model Baumeister and Hamilton (2015c)in order to analyse the impact of oil market shocks on Russian macroeconomic indicators.

5.3 Model

The results of this paper are based on the estimated SBVAR.

The vector of endogenous variables consists of two parts: the variables describing the world oil market and one variable being a Russian economy indicator. The first block of the variables identifies the oil shocks. I then measure how these identified shocks affect a Russian indicator of economic activity included in the model.

For the shocks identification, the model by Baumeister and Hamilton (2015c) is used. The model by Baumeister and Hamilton (2015c) is extended here to estimate the role of the oil market shocks as sources of business cycles. Most of notations here are the same as in Baumeister and Hamilton (2015c). A feature of that model (in contrast to Kilian (2009), Kilian and Murphy (2012), Peersman and Van Robays (2009)) is including stocks in the vector of endogenous variables. It permits to obtain more reasonable estimates of demand and supply elasticities in the oil market. At the same time, using a variable that can be estimated with a measurement error only makes the estimation more tricky and requires some changes of the estimation algorithm with respect to the baseline model.

The oil stocks change if the quantities supplied and demanded are not equal:

$$Q_t^S - Q_t^D = \Delta I_t^*, \tag{5.1}$$

where Q_t^D is a quantity demanded of oil in period t, Q_t^S is a quantity of oil produced in period t, and ΔI_t^* is growth of world oil stocks. The star as the upper index means that the variable is observed with a measurement error. Let $q = 100 \ln(Q_t/Q_{t-1})$ be an observed monthly growth rate of oil production. Then monthly growth rate of the oil demand is written as $q_t - \Delta i_t^*$, where $\Delta i_t^* = 100\Delta I_t^*/Q_{t-1}$.

The model used for estimation is written as:

$$q_t = \alpha_{qp} p_t + b_1' x_{t-1} + u_{1t}^* \quad (5.2)$$

$$z_t = \alpha_{zp} p_t + b_2' x_{t-1} + u_{2t}^* \quad (5.3)$$

$$q_t = \beta_{qz} z_t + \beta_{qp} p_t + \chi^{-1} \Delta i_t + b_3' x_{t-1} + u_{3t}^* - \chi^{-1} e_t \quad (5.4)$$

$$\Delta i_t = \psi_1 q_t + \psi_2 z_t + \psi_3 p_t + b_4' x_{t-1} + \chi u_{4t}^* + e_t \quad (5.5)$$

$$\upsilon_t = \gamma_1 q_t + \gamma_2 z_t + \gamma_3 p_t + \gamma_4^* \Delta i_t^* + b_5' x_{t-1} + u_{5t}^*, \quad (5.6)$$

where q_t is the world monthly oil production rate, z_t is the world economic activity index, p_t is the monthly oil price growth rate, Δi_t is the measure of world stocks growth rate, υ_t is one of the main indicators of the Russian economy, the effect of oil market shocks on υ_t is measured, x_{t-1} represents all lags of endogenous variables: $(x_{t-1}' = (y_{t-1}', y_{t-2}', \ldots, y_{t-m}', 1)')$ and $y_t = (q_t, z_t, p_t, \Delta i_t, \upsilon_t)'$, u_{1t}^* is the oil supply shock, u_{2t}^*- economic activity shock, u_{3t}^* is the oil market specific demand shock, u_{4t}^* is the stocks demand shock that is often titled as speculative demand shock, u_{5t}^* is a internal economy non-oil market shock. Equation (5.2) is the oil supply curve. Equation (5.3) describes the economic activity factors. Equation (5.4) represents the inverse demand function, equation (5.5) describes the stocks dynamics and equation (5.6) describes the dynamics of the Russian times series. In the paper, three different Russian variables are used. They are real monetary income, industrial production, and CPI inflation.

Two last equations in the system above are written under assumption that only a part of the world oil shocks is measured:

$$\Delta i_t = \chi \Delta i_t^* + e_t, \quad (5.7)$$

where $\chi < 1$ is a parameter that represents the measurable part of world oil stocks.

5.4 Estimation

The system written above may be written in the matrix form as:

$$\tilde{A} y_t = \tilde{B} x_{t-1} + \tilde{u}_t, \quad (5.8)$$

$$y_t = (q_t, z_t, p_t, \Delta i_t, \upsilon_t)' \quad (5.9)$$

where the dimensions of the matrix $\tilde{A}$ are (5×5), and the matrix can be written as:

$$\tilde{A} = \begin{bmatrix} 1 & 0 & -\alpha_{qp} & 0 & 0 \\ 0 & 1 & -\alpha_{zp} & 0 & 0 \\ 1 & -\beta_{qz} & -\beta qp & -\chi^{-1} & 0 \\ -\psi_1 & -\psi_2 & -\psi 3 & 1 & 0 \\ -\gamma_1 & -\gamma_2^* & -\gamma 3 & -\gamma 4 & 1 \end{bmatrix} \tag{5.10}$$

and the shocks vector is as follows:

$$\tilde{u}_t = \begin{bmatrix} u_{1t}^* \\ u_{2t}^* \\ u_{3t}^* - \chi^{-1} e_t \\ \chi u_{4t}^* + e_t \\ u_{5t}^* - \gamma_4 e_t \end{bmatrix} \tag{5.11}$$

Due to the measurement error taken into account explicitly, the covariance matrix $D^* = cov(u_{it}, u_{jt})$ is not diagonal:

$$D^* = \begin{bmatrix} d_{11}^* & 0 & 0 & 0 & 0 \\ 0 & d_{22}^* & 0 & 0 & 0 \\ 0 & 0 & d_{33}^* + \chi^{-2}\sigma_e^2 & -\chi^{-1}\sigma_e^2 & \gamma_4 \chi^{-1}\sigma_e^2 \\ 0 & 0 & -\chi^{-1}\sigma_e^2 & d_{44}^*\chi^{-2} + \sigma_e^2 & -\gamma_t \sigma_e^2 \\ 0 & 0 & \gamma_4 \chi^{-1}\sigma_e^2 & -\gamma_t \sigma_e^2 & d_{55}^* + \gamma_4^2 \sigma_e^2 \end{bmatrix} \tag{5.12}$$

To rewrite the model in its usual form with a diagonal covariance matrix of structural shocks, it is possible to multiply both sides of the equation (5.8) by an auxiliary matrix Γ given as:

$$\Gamma = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \rho & 1 & 0 \\ 0 & 0 & \phi & \tau & 1, \end{bmatrix} \tag{5.13}$$

where $\rho = -\frac{D_{34}^*}{D_{33}^*}$, $\tau = \frac{D_{34}^* D_{35^*} - D_{33}^* D_{45}^*}{D_{33}^* D_{44}^* - D_{34}^{*2}}$, $\phi = \frac{-D_{53}^* - D_{43}^* \tau}{D_{33}^*}$ and D_{ij}^* defines an element in the row i and column j of the matrix D^*. Defining $A = \Gamma\tilde{A}$, $B = \Gamma\tilde{B}$ and $u_t = \Gamma\tilde{u}_t$ permits to rewrite model (5.8) in the usual SVAR form:

$$Ay_t = Bx_{t-1} + u_t, \quad u_t \sim i.i.dN(0, D) \tag{5.14}$$

with a diagonal matrix $D = \Gamma\tilde{D}\Gamma'$.

The paper is estimated following the explicit sign restrictions algorithm proposed by Baumeister and Hamilton (2015b). It means that the restrictions are imposed on the parameters of contemporaneous interdependence (matrix A in (5.14)). Some of those parameters, as explained above, have clear economic meaning (as price elasticity of oil demand, for example), and must be either positive or negative. This feature is exploited in the estimation algorithm and priors in the form of the truncated Student distributions are imposed.

Contrary to Baumeister and Hamilton (2015c), in this paper the priors are imposed directly on the parameters of matrix A and not $\tilde{A}$. This is regarded as the only possible solution that makes using posteriors derived by Baumeister and Hamilton (2015b) still possible [1]. For all three datasets, the model is estimated with the RW-MCMC routine using $2 \cdot 10^5$ iterations, half of which is burned-in.

5.5 Empirical results

For all three datasets, impulse response functions, historical decompositions, and forecast error decompositions (FEVD) are calculated. All impulse response functions and historical decomposition of the oil price dynamics are given in the Appendix. The FEVD is presented here.

The Tables 1-3 contain the FEVD at the 12-month horizon. Table 1 shows the decomposition for the set with real money incomes added as an additional Russian variable, Table 2 shows the decomposition for the set with industrial production, and Table 3 shows the decomposition for the set including the CPI inflation.

[1]The problem is due to the measurement error in the stocks variable. It requires pre-multiplying the system by the auxiliary matrix. If prior distributions on the parameters of $\tilde{A}$ are imposed, the prior and posterior distributions refer to different matrices. This clearly contradicts the algorithm proposed by Baumeister and Hamilton (2015b).

	[1]	[2]	[3]	[4]	[5]
Oil supply shock	60	4	38	5	4
Economic activity shock	7	83	12	5	3
Consumption demand shock	12	4	21	60	3
Inventory demand	19	5	24	28	4
Internal shock	2	4	5	2	86

Table 5.1: FEVD for a dataset with real monetary incomes in Russia, [1] - Oil production; [2] - World industrial production; [3] - Oil price; [4] - Stocks ; [5] - Real monetary incomes

	[1]	[2]	[3]	[4]	[5]
Oil supply shock	60	5	39	5	7
Economic activity shock	8	84	14	5	16
Consumption demand shock	12	5	22	60	6
Inventory demand	17	4	23	27	5
Internal shock	3	2	2	2	66

Table 5.2: FEVD for a dataset with Russian industrial production index, [1] - [4] as above; [5] - Industrial production

	[1]	[2]	[3]	[4]	[5]
Oil supply shock	58	4	41	5	3
Economic activity shock	8	85	13	5	4
Consumption demand shock	12	4	20	62	5
Inventory demand	20	4	24	25	4
Internal shock	2	2	2	2	85

Table 5.3: FEVD for a dataset with Russian CPI inflation, [1] - [4] - as above; [5] - CPI inflation

Visual analysis of impulse response functions does not give any precise answer about the external shocks effects on real monetary incomes and inflation, as these effects vary from negative to positive on different iterations of the algorithm. However, the economic activity shock and consumption demand shock positively affect the industrial production index (the former shock affects the IP index immediately). According to the median point FEVD estimates, external shocks account for 14% of the mean squared error associated with the 12-month forecast of real incomes and 15 % of inflation with all four shocks explaining approximately the same fraction of the forecast error variance. The estimates show that external shocks account for a surprisingly large share of the MSE associated with a one year forecast of industrial production (about 34). However, the latter result might be upward biased due to the way how the world economic activity index is calculated and needs to be checked with a sensitivity analysis (that will be added shortly). Though the aim of the paper is to determine the role of oil price shocks for the dynamics of some Russian economy macroindicators,as a by-product it is also possible to draw conclusions about the oil price as well.

The FEVD analysis of oil price shows that the oil supply shock affects the price more strongly than other kinds of shocks, though all demand shocks taken together explain the greater part of the MSE than the supply shock does (60% against 40%) at one-year horizon. The historical decomposition graphs demonstrate that the drop of oil price in 2008 is attributed to all four shocks (at the very beginning of that drop, the inventory demand shock was crucial). This conclusion is in line with that of Baumeister and Hamilton (2015c), even though it was drawn on a much longer and less recent sample in their paper. However, unlike Baumeister and Hamilton (2015c), the historical decomposition shows that the economic activity shock did not play a significant role in the oil price decrease in 2014.

5.6 Prior modifications

5.6.1 Modification 1. No effect of the internal shocks on oil market variables

The baseline model laid above implicitly assumes that the covariance matrix $cov(B|A)$ is diagonal and its main diagonal consists of five equivalent blocs. It means that the prior variance of a parameter at a lag value of a variable depends on the number of the lag and the variable itself but not on the equation that contains the variable. For example, the prior variance of a parameter at p_{t-3} is the same in all equations. In many cases this assumption can be considered as plausible if a researcher does not have any special prior information that allows her to believe in different shrinkage in different equations. In this particular model a researcher probably has a such kind of information. Russia is a small open economy and its internal shocks can hardly affect oil market variables even with a lag. It seems plausible that a matrix of prior variances for lag coefficients can be written as follows:

$$V_{ij,p} = \begin{pmatrix} v_{1,1,p} & v_{1,2,p} & v_{1,3,p} & v_{1,4,p} & 0 \\ v_{2,1,p} & v_{2,2,p} & v_{2,3,p} & v_{2,4,p} & 0 \\ v_{3,1,p} & v_{3,2,p} & v_{3,3,p} & v_{3,4,p} & 0 \\ v_{4,1,p} & v_{4,2,p} & v_{4,3,p} & v_{4,4,p} & 0 \\ v_{5,1,p} & v_{5,2,p} & v_{5,3,p} & v_{5,4,p} & v_{5,5,p} \end{pmatrix},$$

where $v_{i,j,p}$ is a prior variance of an element $b_{i,j}$ in a lag matrix at a lag p (B_p).

Taken into account that all parameters $b_{i,5,p}$ have zero prior mean, the modification of the prior covariance matrix implies that any effect of the Russian indicator on oil markets variables is ruled out at any lag. Technically it means that a conjugate Normal - inverted

Wishart distribution for $B|A$ and $D|A, B$ is replaced with independent Normal - inverted Wishart distribution.

Tables 4 - 6 contain the FEVD at a 12-month horizon for this prior distribution. They show the decomposition with the real money income, industrial production and the CPI inflation, respectively, included in the variable set. Zero values in the last rows of all tables is a consequence of the modification of the prior distribution. In all three cases the median ratio of forecast error variance explained by the oil market shocks taken together increases considerably in comparison with the baseline model. According to the median point FEVD, oil market shocks account for 30% of the mean squared error associated with a 12-month forecast of the real money incomes, 45% of industrial production and 25% of inflation.

The impulse response functions show that internal shock does not influence the variables of the identification block (the assumption embedded in the prior) but the effects of oil market shocks on the Russian indicators are close to those revealed in the baseline model. However, we can conclude that the most of the 95% of the IRF show that a negative supply shock and positive demand shock exert a positive effect on Russian real money incomes, at the same time positive inventory demand exerts a negative effect on real monetary incomes. The consumption demand shock and economic activity shock positively affect the industrial production. The negative supply shock probably increases the industrial production but the influence, if there is any, lasts a short period of time. The effects of oil market shocks on CPI inflation in Russia are uncertain as they vary from negative to positive on different iterations of the algorithm.

	[1]	[2]	[3]	[4]	[5]
Oil supply shock	64	4	37	5	8
Economic activity shock	8	87	14	5	4
Consumption demand shock	11	4	23	62	8
Inventory demand	17	4	26	28	11
Internal shock	0	0	0	0	70

Table 5.4: FEVD for a dataset with real monetary incomes in Russia, prior modification 1, [1] - Oil production; [2] - World industrial production; [3] - Oil price; [4] - Stocks ; [5] - Real monetary incomes

5.6.2 Modification 2. Softer prior in oil production and economic activity equations

A second prior modification replaces two exclusion restrictions in the contemporaneous structural parameter matrix A with softer restrictions determined in the form of Student distributions with zero means. This identification scheme implies the contemporaneous interdependence between the economic activity and oil production variables in two first equations of the system.

Therefore, the A-matrix is now written as:

$$\tilde{A} = \begin{bmatrix} 1 & -\alpha_{qy} & -\alpha_{qp} & 0 & 0 \\ -\alpha_{yq} & 1 & -\alpha_{yp} & 0 & 0 \\ 1 & -\beta_{qz} & -\beta qp & -\chi^{-1} & 0 \\ -\psi_1 & -\psi_2 & -\psi 3 & 1 & 0 \\ -\gamma_1 & -\gamma_2^* & -\gamma 3 & -\gamma 4 & 1 \end{bmatrix} \tag{5.15}$$

The main objective of this prior modification is to let data to speak for themselves without imposing restrictions that may not be supported by the data. However, I do not assume the possibility of non-zero parameters at stocks variable in the first two equations for technical reasons. As world stocks are determined with a measurement error, an assumption that the stock variable is included in the first two equations deprives D^* matrix

	[1]	[2]	[3]	[4]	[5]
Oil supply shock	68	4	33	5	8
Economic activity shock	8	86	14	5	19
Consumption demand shock	11	5	27	60	9
Inventory demand	13	4	26	30	9
Internal shock	0	0	0	0	55

Table 5.5: FEVD for a dataset with Russian industrial production index, prior modification 1, [1] - [4] as above; [5] - Industrial production

	[1]	[2]	[3]	[4]	[5]
Oil supply shock	60	5	41	5	5
Economic activity shock	8	87	14	5	5
Consumption demand shock	12	4	21	63	8
Inventory demand	20	4	24	27	8
Internal shock	0	0	0	0	75

Table 5.6: FEVD for a dataset with Russian CPI inflation, prior modification 1, [1] - [4] - as above; [5] - CPI inflation

of all zero elements. Matrix Γ becomes lower triangular and contains 10 parameters. The solution is cumbersome.

Tables 7 - 9 show the FEVD at the 12-month horizon for this alternative prior modification. The tables contain the decomposition with the real money income, industrial production and the CPI inflation, respectively, included in the variable set.

The impulse response functions are similar to those of the baseline model.

	[1]	[2]	[3]	[4]	[5]
Oil supply shock	41	12	51	6	5
Economic activity shock	19	75	19	6	8
Consumption demand shock	15	7	14	64	3
Inventory demand	25	7	17	24	3
Internal shock	0	0	0	0	80

Table 5.7: FEVD for a dataset with real monetary incomes in Russia, prior modifications 1 and 2, [1] - Oil production; [2] - World industrial production; [3] - Oil price; [4] - Stocks ; [5] - Real monetary incomes

5.7 Conclusion

In the paper, I construct an SBVAR model to identify the contribution of oil market structural shocks into some Russian macroeconomic indicators dynamics. The main interest of this paper is the quantitative estimate of the impact of oil market shocks on macroeconomic volatility in Russia. The core of the model used in the paper is the model by Baumeister and Hamilton (2015c) extended by one equation describing the dynamics of a macroeconomic indicator in question. The model is general and can be applied to any economy.

The identified shocks give economically plausible results about their effects on the oil market and global activity, though the sign of some of these effects is predetermined by the sign restrictions embedded in the priors.

The main research question of the paper is answered differently for different indicators. The mean estimate of the forward error variance explained by oil market shocks at a one-year horizon is between 14% and 30% depending on the prior for real monetary incomes and between 15% and 25% for inflation. Conversely, the fraction of the forward error variance of industrial production explained by oil shocks at one year horizon is between 35% and 45% depending on the prior. These values seem to be surprisingly high, even for an oil-dependent economy like Russia's, and need to be examined using a sensitivity

	[1]	[2]	[3]	[4]	[5]
Oil supply shock	29	57	13	5	20
Economic activity shock	24	28	65	7	16
Consumption demand shock	16	7	10	66	5
Inventory demand	31	8	12	22	6
Internal shock	0	0	0	0	53

Table 5.8: FEVD for a dataset with Russian industrial production index, prior modification 1 and 2, [1] - [4] as above; [5] - Industrial production

	[1]	[2]	[3]	[4]	[5]
Oil supply shock	18	9	76	6	6
Economic activity shock	13	82	12	6	6
Consumption demand shock	21	4	7	70	6
Inventory demand	49	5	6	17	4
Internal shock	0	0	0	0	78

Table 5.9: FEVD for a dataset with Russian CPI inflation, prior modification 1 and 2, [1] - [4] - as above; [5] - CPI inflation

analysis.

The empirical results of the estimation allow drawing a conclusion about the oil price dynamics as well. The estimation results show that the oil supply shock affects the price dynamics more strongly than any of the demand shocks. However, all three demand shocks in total drive the oil price dynamics more strongly than the supply shock. As in the paper by Baumeister and Hamilton (2015c), I show here that the oil price drop in 2008 was triggered by all four shocks considered. At the very beginning of that drop, the inventory demand shock played the most significant role. Contrary to their results, the historical variance decomposition of the shocks in this paper shows that the economic activity shock was not a significant driver of the oil price decrease in 2014.

Appendices

.1 Impulse response functions

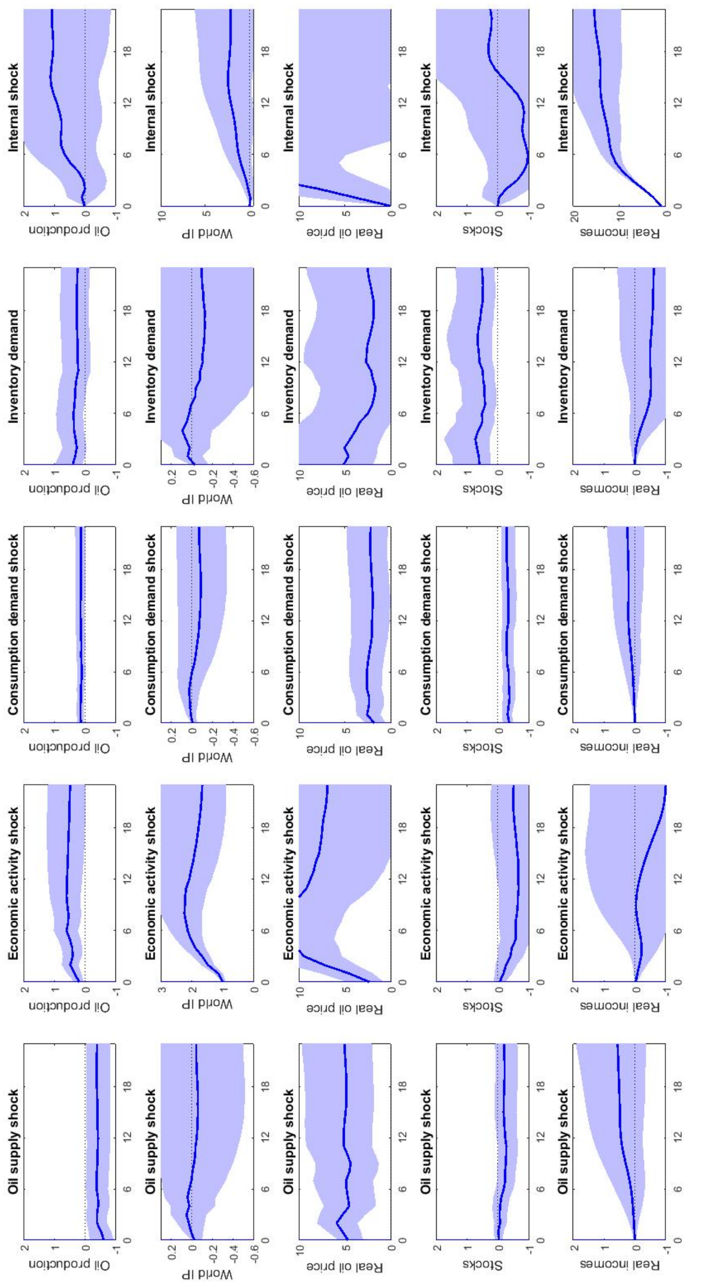


Figure 1: Impulse response functions for all five variables with real money incomes as Russian macroindicator

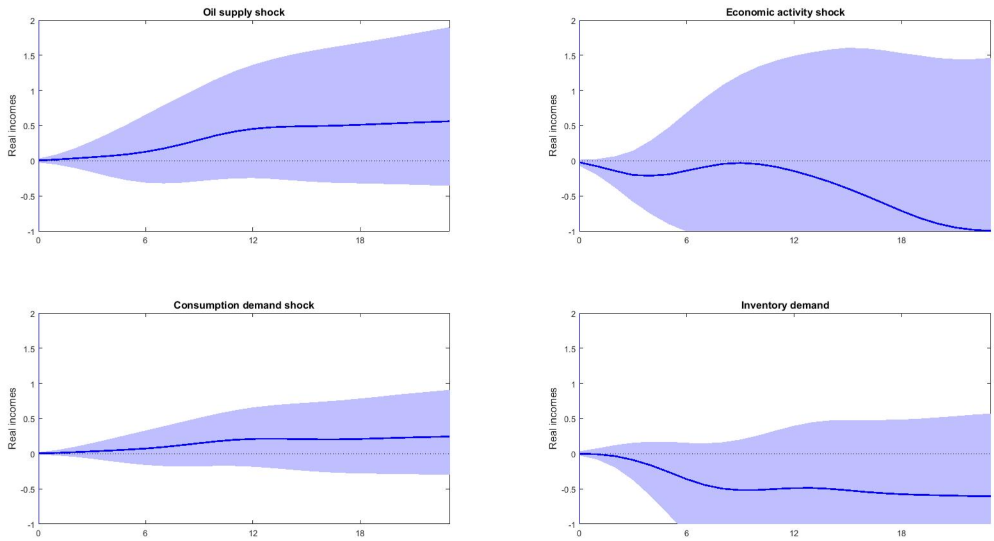


Figure 2: Impulse response functions for real money incomes hit by oil market shocks (the same as shown in last row of Figure 1

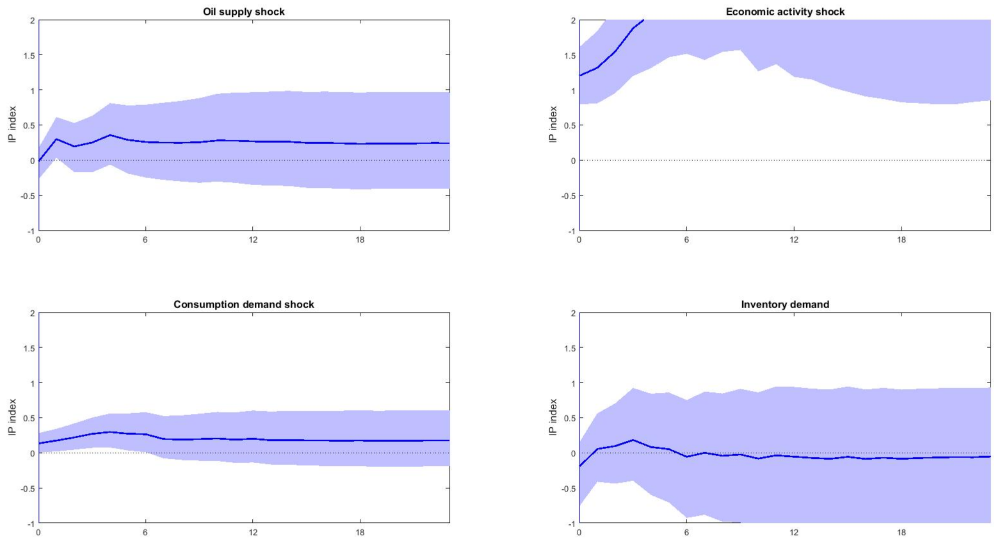


Figure 3: Impulse response functions for industrial production index

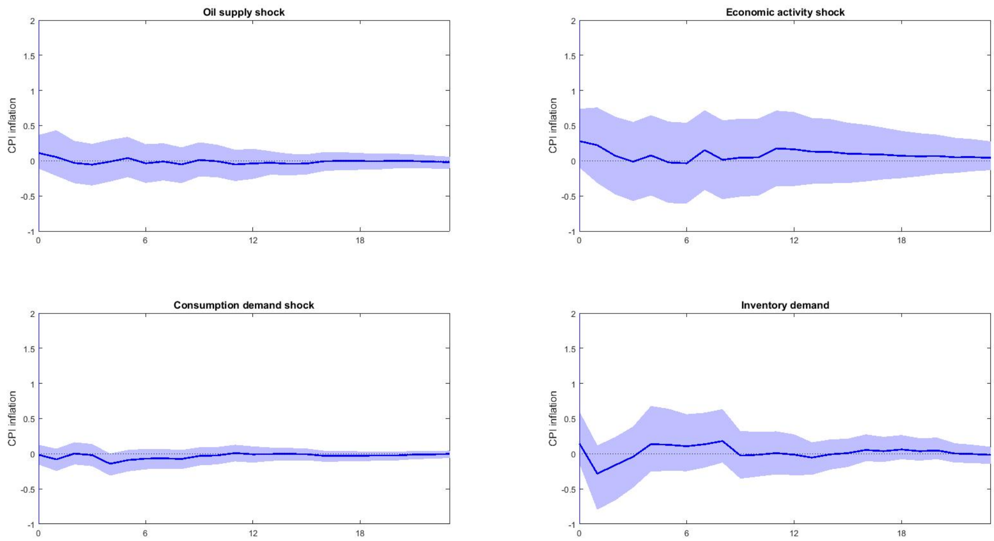


Figure 4: Impulse response functions for CPI inflation

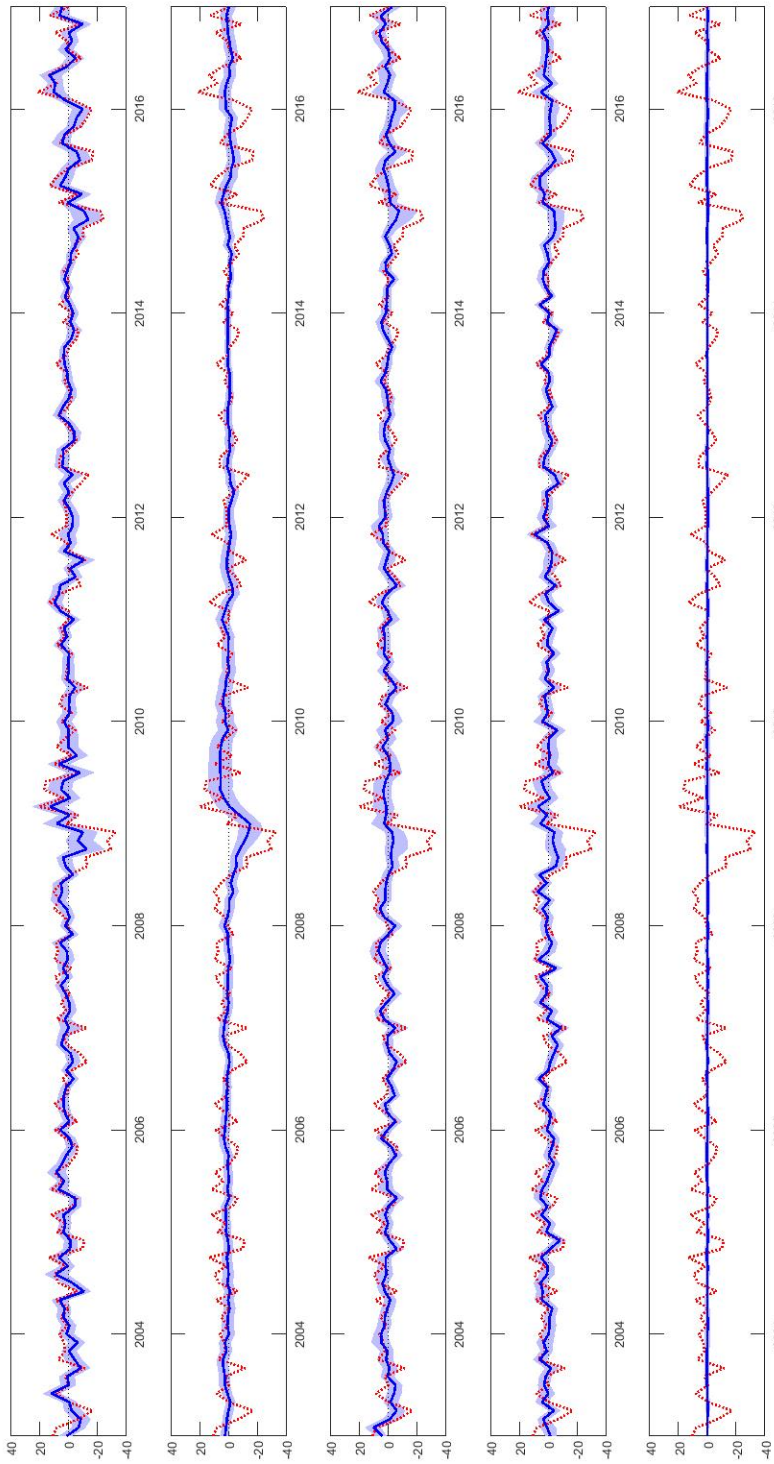


Figure 5: Historical decomposition of oil price dynamics

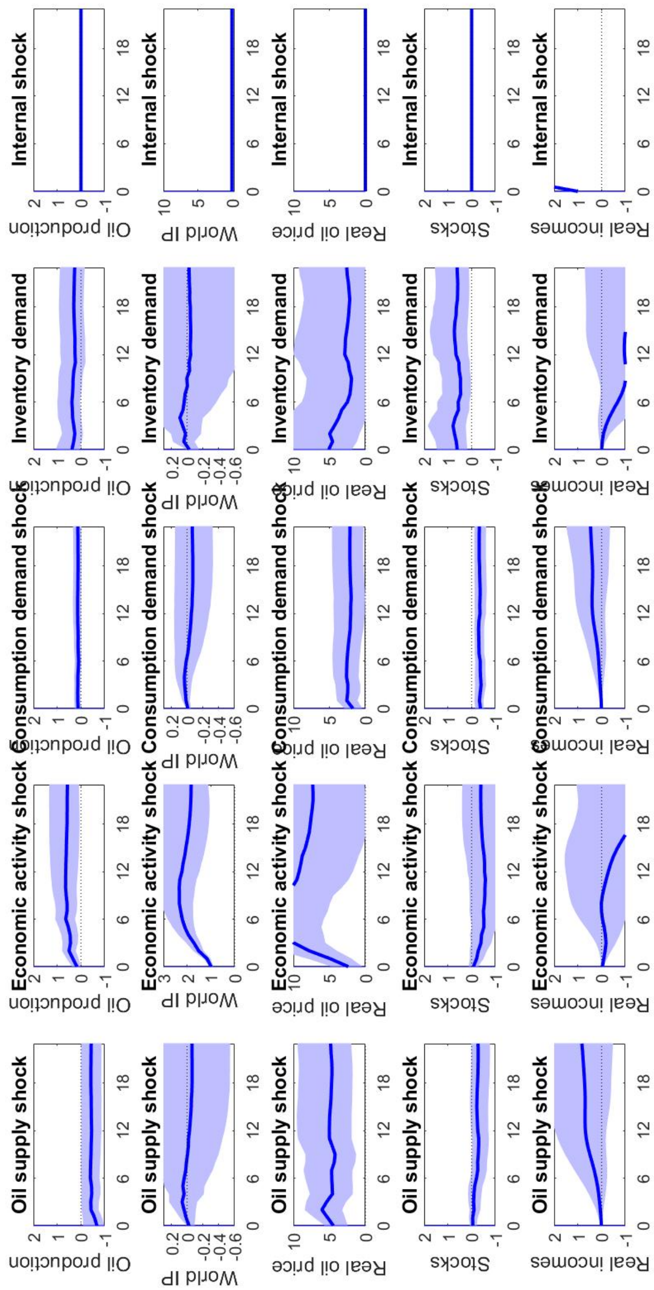


Figure 6: Impulse response functions for all five variables with real money incomes as Russian macroindicator

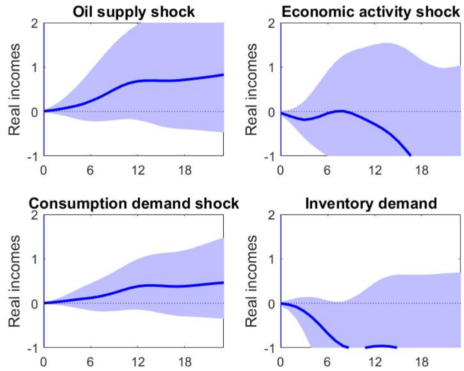


Figure 7: Impulse response functions for real money incomes hit by oil market shocks (the same as shown in last row of Figure 6), prior modification 1

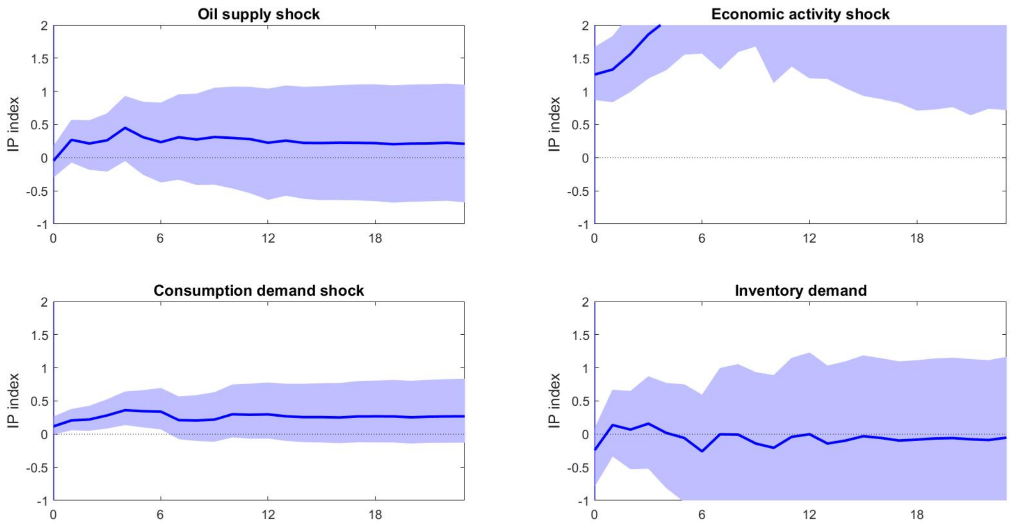


Figure 8: Impulse response functions for industrial production index, prior modification 1

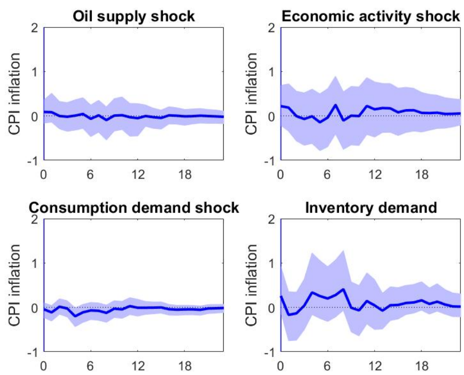


Figure 9: Impulse response functions for CPI inflation, prior modification 1

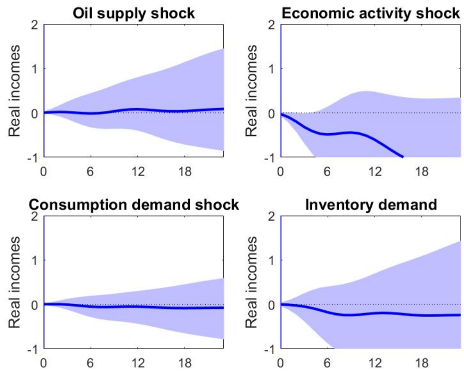


Figure 10: Impulse response functions for real money incomes hit by oil market shocks, prior modification 1 and 2

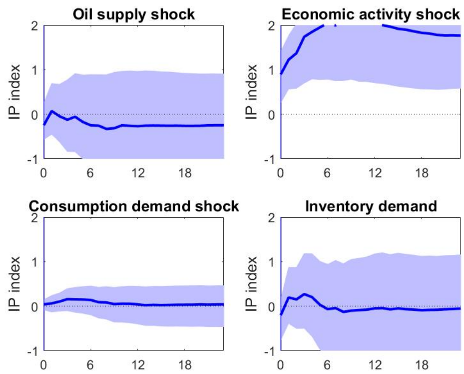


Figure 11: Impulse response functions for industrial production index, prior modification 1 and 2

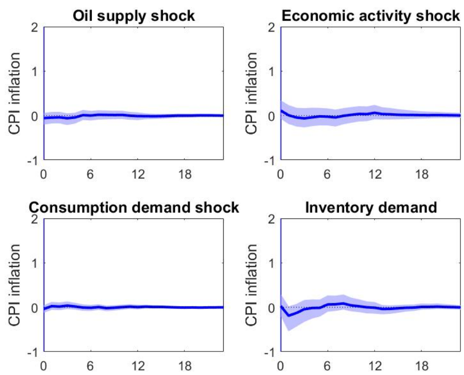


Figure 12: Impulse response functions for CPI inflation, prior modification 1 and 2

General Conclusion

This thesis addressed several research questions related to forecasting and structural modelling of an energy-based economy.

Chapter 1 presented a DSGE model with an economy with windfall revenues from commodity exports. The model parameters were estimated using Russian data but the model is not specific and can be calibrated or estimated for any other commodity-based economy. The goal of the study was twofold. First, we aimed at constructing a model in a general equilibrium framework that reflects some specific features of commodity-exporting countries. Second, a goal was to determine the main shocks responsible for volatility of key macroeconomic variables in Russia and to estimate to what extent the oil price shocks govern the Russian business cycle.

The model yielded plausible estimates, with impulse response functions being in line with economic logic. Historical decomposition showed that the financial crisis of 2009 in Russia was primarily caused by risk premium shock and commodity export shock, which seems reasonable. Forward error variance decomposition showed that non-commodity output both for final and intermediate goods was governed by domestic shocks, while commodity export shock does contribute much to GDP, since GDP explicitly accounts for all export revenues. We conclude that on the span considered in the chapter, the business cycle in Russia is mostly domestically based with a substantial effect of commodity export shock on the GDP dynamics.

Chapter 2 reviewed the algorithms of estimation and forecasting using the reduced-form BVAR models. The chapter presented the array of prior distributions that could be found in applied research and described comprehensively their interdependence. A separate section of this chapter was devoted to the technique of defining the conjugate normal - inverted Wishart distribution with dummy observations. This method is widely used in economic applications but had not been well presented in the reviews of BVAR approach so far. A last part of the chapter discussed how to proceed with point or density forecasting after an estimation of a reduced-form BVAR was completed.

Chapter 3 compared the forecasting accuracy of reduced-form BVAR models of different sizes with the accuracy of frequentist VAR models and an array of univariate models. The forecasting exercise was done for a sample of 23 monthly Russian macroeconomic indicators, and five forecasting horizons that varied from 1 to 12 months. We concluded that for most of the variables and forecasting horizons BVAR models outperform the competitors. We could not confirm the results of some other studies where similar tools were applied to data from developed countries, that higher dimensional models permit to obtain more accurate forecasts. The Russian data showed that a 23-variable BVAR was the most accurate model in only a half cases where a BVAR model performed better than its simpler competitors. For the rest of those cases, a 23-variable BVAR was outperformed by a smaller BVAR model. A robustness check analysis fulfilled with MCS procedure confirmed that the BVAR models were included in the confidence set more often than VARs, but for majority of variables and forecasting horizons the random walk model was not inferior with respect to the BVAR.

Chapter 4 compared the forecasting accuracy of the DSGE model with that of non-structural VAR and BVAR models. The models were applied to Russian data, and the DSGE model accounted for the oil-export revenues explicitly. The DSGE model was the same as constructed in the first chapter of the thesis. In the the BVAR model, the parameters were shrunk with conjugate normal - inverse Wishart prior with modifications (the prior discussed in the second chapter and applied to forecasting of monthly time series in the third chapter of the thesis). Therefore, the forth chapter extended all previous chapters of this dissertation. A characteristic feature the estimation and forecasting procedure is the iterative trend used for all three models to mimic a real-time forecasting exercise. In chapter 4, we conclude that the DSGE models were generally outperformed by the BVAR model. However, the DSGE-forecasts are the most accurate for some variables and some forecasting horizons. We also provide a data-related explanation of the relative failure to forecast some variables in the sample.

Chapter 5 constructs an SBVAR model to quantify the contribution of oil market shocks on macroeconomic volatility. For shocks identification, explicit sign restrictions were used, with prior distributions imposed directly on the matrix of simultaneous dependencies. The model was estimated on Russian data but being general, it can be applied to any economy.

The effects of the identified shocks on oil market market and global activity are in line with economic logic, though the sign of some of them is a simple reflection of restrictions implied by the priors.

The results of the estimation are mixed about the impact of oil price shocks on the real monetary incomes and CPI inflation. However they indicate a strong influence of two demand shocks on the industrial production index. The quantitative estimates measured with forward error variance decomposition are surprisingly high, even for an oil-dependent economy like Russia's, and may be questioned again in future research.

References

Aastveit, K. A. (2014). Oil price shocks in a data-rich environment. *Energy Economics*, *45*(100), 268–279.

Abel, A. (1990). Asset prices under habit formation and catching up with the joneses. *American Economic Review*, *80*(2), 38–42.

Adolfson, M., Lindé, J., & Villani, M. (2007). Forecasting performance of an open economy dsge model. *Econometric Reviews*, *26*(2-4), 289–328.

Ahmad, Y., Lo, M., & Mykhaylova, O. (2011). *Persistence and non-linearity of simulated dsge real exchange rates* (Working Paper No. 11-01). University of Wisconsin.

Alessandri, P. & Mumtaz, H. (2014). *Financial conditions and density forecasts for us output and inflation*. Working Paper, School of Economics and Finance, Queen Mary, University of London.

Alvarez, I. (2014). Bayesian inference for a covariance matrix. *arXiv preprint arXiv:1408.4050*.

Amir Ahmadi, P. & Uhlig, H. (2015). *Sign restrictions in bayesian favars with an application to monetary policy shocks* (NBER Working Papers No. 21738). National Bureau of Economic Research, Inc.

Amisano, G. & Giacomini, R. (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics*, *25*, 177–190.

Balassa, B. (1964). The purchasing power doctrine: A reappraisal. *Journal of Political Economy*, *72*(6), 584–596.

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large bayesian vector auto regressions. *Journal of Applied Econometrics*, *25*(1), 71–92.

Barsky, R. & Kilian, L. (2002). Do we really know that oil caused the great stagflation? a monetary alternative. In *Nber macroeconomics annual 2001, volume 16* (pp. 137–198). National Bureau of Economic Research, Inc.

Barsky, R. & Kilian, L. (2004). Oil and the macroeconomy since the 1970s. *Journal of Economic Perspectives*, *18*(4), 115–134.

Batte, L., Bénassy-Quéré, A., Carton, B., & Dufrénot, G. (2009). *Term of trade shocks in a monetary union: An application to west africa* (Working Paper No. 2009-07). CEPII.

Baumeister, C. & Hamilton, J. (2015a). *Optimal inference about impulse-response functions and historical decompositions in incompletely identified structural vector autoregressions.*

Baumeister, C. & Hamilton, J. (2015b). Sign restrictions, structural vector autoregressions, and useful prior information. *Econometrica*, *83*(5), 1963–1999.

Baumeister, C. & Hamilton, J. (2015c). *Structural interpretation ov vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks.*

Baumeister, C. & Peersman, G. (2013). Time-varying effects of oil supply shocks on the us economy. *American Economic Journal: Macroeconomics*, *5*(4), 1–28.

Baumeister, C., Peersman, G., & Van Robays, I. (2010). The economic consequences of oil shocks: Differences across countries and time. In R. Fry, C. Jones, & C. Kent (Eds.), *Inflation in an era of relative price shocks*. Reserve Bank of Australia.

Bauwens, L., Lubrano, M., & Richard, J.-F. (2000). *Bayesian inference in dynamic econometric models*. Oxford University Press.

Beauchemin, K. & Zaman, S. (2011). *A medium scale forecasting model for monetary policy* (Working Paper No. 1128). Federal Reserve Bank of Cleveland.

Beckner, S. K. (1996). *Back from the brink: The greenspan years*. John Wiley & Sons Inc.

Benedictow, A., Fjaertoft, D., & Lofsnaes, O. (2013). Oil dependency of the russian economy: An econometric analysis. *Economic Modelling*, *32*(100), 400–428.

Berg, T. O. & Henzel, S. (2013). *Point and density forecasts for the euro area using many predictors: Are large bvars really superior?* ifo Working Paper.

Bergin, P. & Feenstra, R. (2001). Pricing-to-market, staggered contracts and real exchange rate persistence. *Journal of International Economics*, *54*(2), 333–359.

Bernanke, B. & Boivin, J. (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics*, *50*(3), 525–546.

Bernanke, B., Boivin, J., & Eliasz, P. S. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (favar) approach. *The Quarterly Journal of Economics*, *120*(1), 387–422.

Bjørnland, H. (2008). Monetary policy and exchange rate interactions in a small open economy. *Scandinavian Journal of Economics*, *110*(1), 197–221.

Blake, A. & Mumtaz, H. (2012). *Applied bayesian econometrics for central bankers* (1st ed.). Centre for Central Banking Studies, Bank of England.

Blanchard, O. & Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances. *American Economic Review*, *79*(4), 655–73.

Blanchard, O. & Gali, J. (2008). *The macroeconomic effects of oil shocks: Why are the 2000s so different from the 1970s?* (CEPR Discussion Papers No. 6631). C.E.P.R. Discussion Papers.

Bloor, C. & Matheson, T. (2010). Analysing shock transmission in a data-rich environment: A large bvar for new zealand. *Empirical Economics*, *39*(2), 537–558.

Bruno, M. & Sachs, J. D. (1985). *Economics of worldwide stagflation*. National Bureau of Economic Research, Inc.

Calvo, G. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics*, *12*(3), 383–398.

Canova, F. (2007). *Methods for applied macroeconomic research*. Princeton University Press.

Carriero, A., Clark, T. E., & Marcellino, M. (2015). Bayesian vars: Specification choices and forecast accuracy. *Journal of Applied Econometrics*, *30*(1), 46–73.

Carriero, A., Kapetanios, G., & Marcellino, M. (2012). Forecasting government bond yields with large bayesian vector autoregressions. *Journal of Banking & Finance*, *36*(7), 2026–2047.

Cavalcanti, T. & Jalles, J. (2013). Macroeconomic effects of oil price shocks in brazil and in the united states. *Applied Energy*, *104*(100), 475–486.

Cespedes, L. F., Chang, R., & Velasco, A. (2004). Balance sheets and exchange rate policy. *American Economic Review*, *94*(4), 1183–1193.

Chari, V., Kehoe, P., & E.McGrattan. (2002). Can sticky price models generate volatile and persistent real exchange rate. *Review of Economic Studies*, *69*(3), 533–563.

Charnavoki, V. & Dolado, J. (2014). The effects of global shocks on small commodity-exporting economies: Lessons from canada. *American Economic Journal: Macroeconomics*, *6*(2), 207–37.

Christiano, L., Eichembaum, M., & Evans, C. (2005). Nominal rigidities and the dynamic effects to a shock of monetary policy. *Journal of Political Economy*, *113*(1), 1–45.

Christoffel, K., Coenen, G., & Warne, A. (2008). *Tne new area-wide model of the euro-area. a micro-founded open-eonomy model for forecasting and policy analysis* (Working paper No. 944). European Central Bank.

Christoffel, K., Warne, A., & Coenen, G. (2010). *Forecasting with dsge models* (Working Paper Series No. 1185). European Central Bank.

Corsetti, G., Dedola, L., & Leduc, S. (2008). High exchange volatility and low pass-through. *Journal of Monetary Economics*, *55*(6), 1113–1128.

Cuñado, J. & Pérez de Gracia, F. (2003). Do oil price shocks matter? evidence for some european countries. *Energy Economics*, *25*(2), 137–154.

Curdia, V. & Finocchiaro, D. (2005). *An estimates dsge model for sweden with a monetary regime change* (Seminar Paper No. 740). Institute for International Economic Studies. Institute for International Economic Studies Seminar Paper N 740.

Dam, N. & Linaa, J. (2005). *Assessing the welare cost of a fixed exchange rate policy* (Working Paper No. 2005-4). University of Copenhagen.

De Mol, C., Giannone, D., & Reichlin, L. (2008). Forecasting using a large number of predictors: Is bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, *146*(2), 318–328.

Dees, S. & Güntner, J. (2014). *Analysing and forecasting price dynamics across euro area countries and sectors: A panel var approach* (Working Paper Series No. 1724). European Central Bank.

Del Negro, M. & Schorfheide, F. (2011). Bayesian macroeconometrics. *The Oxford handbook of Bayesian econometrics*, 293–389.

Demeshev, B. & Malakhovskaya, O. (2016). Bvar mapping. *Applied Econometrics*, *43*, 118–141.

Dib, A. (2008). *Welfare effects of commodity price and exchange rate volatilities in a multi-sector small open economy model.* Bank of Canada.

Dib, A. (2011). Monetary policy in estimated models of small open and closed economies. *Open Economies Review*, *22*(5), 769–796.

Diebold, F., Gunther, T. A., & Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, *39*(4), 863–83.

Diebold, F. & Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–63.

Doan, T., Litterman, & Sims. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric reviews*, *3*(1), 1–100.

Edelstein, P. & Kilian, L. (2007). The response of business fixed investment to changes in energy prices: A test of some hypotheses about the transmission of energy price shocks. *The B.E. Journal of Macroeconomics*, *7*(1), 1–41.

Edge, R. M. & Gürkaynak, R. (2011). *How useful are estimated dsge model forecasts?* (Finance and Economics Discussion Series No. 2011-11). Board of Governors of the Federal Reserve System (US).

Erceg, C., Henderson, D., & Levin, A. (2000). Optimal monetary policy with staggered wage and price contracts. *Journal of Monetary Economics*, *46*(2), 281–313.

Faia, E. & Monacelli, T. (2008). Optimal monetary policy in a small open economy with home bias. *Journal of Money, Credit and Banking*, *40*(4), 721–750.

Favero, C. A. (2001). *Applied macroeconometrics*. Oxford University Press.

Fernandez-Villaverde, J. (2010). The econometrics of dsge models. *SERIEs*, *1*(1), 3–49.

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and statistics*, *82*(4), 540–554.

Gali, J. (2008). *Monetary policy, inflation and the business cycle: An introduction to the new keynesian framework*. Princeton University Press.

Gerdrup, K. R., Jore, A. S., Smith, C., & Thorsrud, L. (2009). *Evaluating ensemble density combination - forecasting gdp and inflation* (Working Paper No. 2009/19). Norges Bank.

Geweke, J. (1977). The dynamic factor analysis of economic time series. In D. Aigner & A. Goldberger (Eds.), *Latent variables in socio-economic models 1*. North-Holland, Amsterdam.

Geweke, J. & Amisano, G. (2010). Comparing and evaluating bayesian predictive distributions of asset returns. *International Journal of Forecasting*, *26*(2), 216–230.

Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society Series B*, *69*(2), 243–268.

Gonzalez-Rivera, G. & Sun, Y. (2015). Generalized autocontours: Evaluation of multivariate density models. *International Journal of Forecasting*, *31*(3), 799–814.

Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society Ser. B,*

Guerron, P. & Nason, J. (2013). Bayesian estimation of dsge models. In *Handbook of research methods and applications in empirical macroeconomics* (Chap. 21, pp. 486–512). Edward Elgar Publishing.

Hamilton, J. (1983). Oil and the macroeconomy since world war ii. *Journal of Political Economy*, *91*(2), 228–48.

Hamilton, J. (1996). This is what happened to the oil price-macroeconomy relationship. *Journal of Monetary Economics*, *38*(2), 215–220.

Hansen, P. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, *23*, 365–380.

Hansen, P., Lunde, A., & Nason, J. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, *13*(2), 281–291.

He, Y., Wang, S., & Lai, K. K. (2010). Global economic activity and crude oil prices: A cointegration analysis. *Energy Economics*, *32*(4), 868–876.

Herrera, A. M. & Pesavento, E. (2009). Oil price shocks, systematic monetary policy, and the "great moderation". *Macroeconomic Dynamics*, *13*(01), 107–137.

Hooker, M. A. (1996). What happened to the oil price-macroeconomy relationship? *Journal of Monetary Economics*, *38*(2), 195–213.

Hyndman, R. (2019). *Package 'forecast'*. Retrieved from https://cran.r-project.org/web/packages/forecast/forecast.pdf

Ivashchenko, S. (2013). Dynamic stochastic general equilibrium model with banks and endogenous defaults of firms. *Journal of the New Economic Association*, *19*(3), 27–50.

Jin, G. (2008). The impact of oil price shock and exchange rate volatility on economic growth : A comparative analysis for russia, japan and china. *Research Journal of International Studies*, *8*, 98–111.

Justiniano, A. & Preston, B. (2010). Monetary policy and uncertainty in an empirical small open-economy model. *Journal of Applied Econometrics*, *25*(1), 93–128.

Kadiyala, K. & Karlsson, S. (1997). Numerical methods for estimation and inference in baesian var-models. *Journal of Applied Econometrics*, *12*(2), 99–132.

Karlsson, S. (2012). *Forecasting with bayesian vector autoregressions* (Working Papers No. 2012:12). Örebro University, School of Business.

Karlsson, S. (2013). Forecasting with bayesian vector autoregression. (Chap. Chapter 15, Vol. 2, pp. 791–897). Elsevier.

Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, *99*(3), 1053–69.

Kilian, L. & Murphy, D. (2012). Why agnostic sign restrictions are not enough: Understanding the dynamics of oil market var models. *Journal of the European Economic Association*, *10*(5), 1166–1188.

Kim, S. & Roubini, N. (2000). Exchange rate anomalies in the industrial countries: A solution with a structural var approach. *Journal of Monetary Economics*, *45*(3), 561–586.

Knetter, M. (1993). International comparisons of pricing-to-market behavior. *American Economic Review*, *83*(3), 473–486.

Kollmann, R. (2001). The exchange rate in a dynamic-optimizing business cycle model with nominal rigidities: A quantitative investigation. *Journal of International Economics*, *55*(2), 243–262.

Koop, G. (2013). Forecasting with medium and large bayesian vars. *Journal of Applied Econometrics*, *28*(2), 177–203.

Koop, G. & Korobilis, D. (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends(R) in Econometrics*, *3*(4), 267–358.

Kreptsev D.A., S. S. (2016). *Dsge-models of russian economy with a small number of equations*. Bank of Russia.

Kwiatkowski, D., Phillips, P., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, *54*(1-3), 159–178.

Lane, P. & Milesi-Ferretti, G. M. (2001). *Long-term capital movements* (NBER Working Papers No. 8366). National Bureau of Economic Research, Inc.

Lindé, J., Nessen, M., & Soderstrom, U. (2009). Monetary policy in an estimated open-economy model with imperfect pass-through. *International Journal of Finance and Economics*, *14*(4), 301–333.

Lippi, F. & Nobili, A. (2012). Oil and the macroeconomy: A quantitative structural analysis. *Journal of the European Economic Association*, *10*(5), 1059–1083.

Litterman. (1979). *Techniques of forecasting using vector autoregressions* (Working Papers No. 115). Federal Reserve Bank of Minneapolis.

Litterman. (1986). Forecasting with bayesian vector autoregressions—five years of experience. *Journal of Business & Economic Statistics*, *4*(1), 25–38.

Lubik, T. & Schorfheide, F. (2007). Do central banks respond to exchange rate movements? a structural investigation. *Journal of Monetary Economics*, *54*(4), 1069–1087.

Malakhovskaya, O. (2013). *Exchange rate policy in countries with imperfect financial markets under conditions of negative balance-of-payment shocks*. Laboratory for Macroeconomic Analysis (in Russian).

Malakhovskaya, O. & Pekarsky, S. (2012). Research on causality interdependencies in macroeconomic: Nobel prize of 2011 (in russian). *Economic Journal of Higher School of Economics*, *16*(1), 3–30.

Malakhovskaya, O. (2016). Dsge-based forecasting: What should our perspective be? *VOPROSY ECONOMIKI*, *12*.

Malakhovskaya, O. & Minabutdinov, A. (2014). Are commodity price shocks important? a bayesian estimation of a dsge model for russia. *International Journal of Computational Economics and Econometrics*, *4*(1/2), 148–180.

Mariano, R. & Preve, D. (2012). Statistical tests for multiple forecast comparison. *Journal of Econometrics*, *169*(1), 123–130.

Meese, R. & Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, *14*(1-2), 3–24.

Melolinna, M. (2012). *Macroeconomic shocks in an oil market var* (Working Paper Series No. 1432). European Central Bank.

Mitchell, J. & Wallis, K. (2011). Evaluating density forecasts: Forecast combinations, model mixtures, calibration and sharpness. *Journal of Applied Econometrics*, *26*(6), 1023–1040.

Mork, K. A., Olsen, O., & Mysen, H. T. (1994). Macroeconomic responses to oil price increases and decreases in seven oecd countries. *The Energy Journal*, *Volume15*(Number 4), 19–36.

Peersman, G. & Van Robays, I. (2009). Oil and the euro area economy. *Economic Policy*, *24*, 603–651.

Pesaran, M., Schuermann, T., & Weiner, S. (2004). Modeling regional interdependencies using a global error-correcting macroeconometric model. *Journal of Business & Economic Statistics*, *22*, 129–162.

Peterman, W. (2012). *Reconciling micro and macro estimates of the frisch labor supply elasticity* (Finance and Economics Discussion Series No. 2012-75). Board of Governors of the Federal Reserve System (U.S.)

Polbin, A. (2014). Econometric estimation of a structural macroeconomic model for the russian economy. *Applied Econometrics*, *33*(1), 3–29.

Primiceri, G. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, *72*(3), 821–852.

Rautava, J. (2002). *The role of oil prices and the real exchange rate on russia's economy.* BOFIT Discussion Paper №3.

Robertson, J. C. & Tallman, E. W. (1999). Vector autoregressions: Forecasting and reality. *Economic Review-Federal Reserve Bank of Atlanta*, *84*(1), 4.

Romano, J. P. & Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, *73*(4), 1237–1282.

Schmitt-Grohe, S. & Uribe, M. (2003). Closing small open economy models. *Journal of International Economics*, *61*(1), 163–185.

Scholl, A. & Uhlig, H. (2008). New evidence on the puzzles: Results from agnostic identification on monetary policy and exchange rates. *Journal of International Economics*, *76*(1), 1–13.

Semko, R. (2013). Optimal economic policy and oil shocks in russia. *Ekonomska istraživanja – Economic Research*, *26(2)*, 364–379.

Shulgin, A. (2014). How much monetary policy rules do we need to estimate dsge model for russia? *Applied Econometrics*, *36*(4), 3–31.

Sims. (1980). Macroeconomics and reality. *Econometrica*, *48*(1), 1–48.

Sims. (1992). Bayesian inference for multivariate time series with trend. In *American statistical association meetings.*

Sims. (1993). A nine-variable probabilistic macroeconomic forecasting model. In *Business cycles, indicators and forecasting* (pp. 179–212). University of Chicago Press.

Sims, C. (2003). *Probability models for monetary policy decisions*. Retrieved from http://sims.princeton.edu/yftp/Ottawa/ProbModels.pdf

Sims, C. A. & Zha, T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, 949–968.

Smets, F. & Wouters, R. (2003). An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association*, *1*(3), 1123–1175.

Smets, F. & Wouters, R. (2007). *Shocks and frictions in us business cycles: A bayesian dsge approach* (Working paper No. 722). European Central Bank.

Sosunov, K. & Zamulin, O. (2007). *Monetary policy in an economy sick with dutch disease* (Working Paper No. 2007-07). Laboratory for Macroeconomic Analysis.

Stock, J. & Watson, M. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–62.

Stock, J. & Watson, M. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of macroeconomics* (Chap. Chapter 8, Vol. 2, pp. 415–525). Elsevier.

Taro, I. (2010). Interest rate rule for the russian monetary policy: Nonlinearity and asymmetricity. *Hitotsubashi Journal of Economics*, *51*(1), 1–11.

Taylor, A. & Taylor, M. (2004). The purchasing power parity debate. *Journal of Economic Perspectives*, *18*, 135–158.

Tsyplakov, A. (2013). *Evaluation of probabilistic forecasts: Proper scoring rules and moments* (MPRA Paper No. 45186). University Library of Munich, Germany.

Uhlig, H. (1997). Bayesian vector autoregressions with stochastic volatility. *Econometrica*, *65*(1), 59–74.

Vdovichenko, A. & Voronina, V. (2006). Monetary policy rules and their application in russia. *Research in International Business and Finance*, *20*, 145–162.

White, H. (2000). A reality check for data snooping. *Econometrica*, *68*(5), 1097–1126.

Woodford, M. (1996). *Interest and prices: Foundations of a theory of monetary policy*. Princeton University Press.

Yudaeva, K., Ivanova, N., & Kamenskikh, M. (2010). *What does the bank of russia target?* Center for Macroeconomic Research, Sberbank.

Zellner, A. (1996). *An introduction to bayesian inference in econometrics*. Wiley Classics Library. Wiley.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences de l'homme et de la société (SHS)

Titre : Essais sur la prévision et modélisation d'une économie riche en ressources pétrolières

Mots clés : prévision, DSGE, BVAR, SBVAR

Résumé : Il y a un consensus que la sévérité des chocs sur les marchés pétroliers tend à diminuer, ainsi que la dépendance des économies développées vis-à-vis de ces chocs. Les pays développés sont généralement les importateurs d'énergie et l'effet des chocs pétroliers sur les pays exportateurs de pétrole peut être différent, surtout s'il s'agit des pays dont la grande partie de l'exportation est le pétrole ou les produits pétroliers. En outre, l'orientation sur l'exportation des matières premières peut modifier la performance relative des modèles économétriques qui sont généralement utilisés pour les prévisions. La thèse étudie et développe des modèles de l'analyse structurelle et de la prévision à court terme d'une économie exportatrice de pétrole où les données russes sont utilisées pour toutes les applications empiriques. Le premier chapitre est consacré à la construction d'un modèle DSGE pour un pays exportateur de matières premières. Le modèle DSGE est estimé par des méthodes bayésiennes. Nous constatons qu'en dépit de l'impact important sur le PIB des chocs pétroliers, les cycles économiques en Russie sont essentiellement d'origine intérieure. Le deuxième chapitre examine comment les méthodes bayésiennes peuvent être appliquées aux prévisions à l'aide d'un modèle BVAR. Le troisième chapitre applique ces techniques et compare la performance d'un groupe de modèles non structurels (univariés et multivariés) pour prévoir un ensemble d'indicateurs macroéconomiques russes. Dans le quatrième chapitre, les prévisions se sont concentrées sur les modèles structurels multivariés (DSGE) et non structurels (BVAR). Le cinquième chapitre quantifie l'effet de différents types de chocs pétroliers sur plusieurs variables macroéconomiques russes.

Title : Essays on forecasting and modelling of an energy-based economy

Keywords : forecasting, DSGE, BVAR, SBVAR

Abstract : It is generally agreed that the severity of oil markets shocks tends to decrease as does dependence of developed economies on those shocks. Developed countries are generally energy importers, and the effect of oil market shocks on oil-exporting countries may be different, especially if energy represents a large percentage of the country's exports. In addition, the focus on commodity exports may change the relative forecasting performance of econometric models that are generally used for forecasting. This thesis studies and develops models for structural analysis and short-term forecasting of an oil-exporting economy using Russian data for all empirical applications. The first chapter is devoted to a construction of a DSGE model for a country with commodity exports. The DSGE model is estimated by Bayesian methods We find that despite a strong impact of commodity export shocks on GDP, the business cycles in Russia are mostly domestically based. The second chapter discusses how the Bayesian methods may be applied for forecasting with a BVAR model. The third chapter applies these techniques and compares the performance of a group of non-structural models – univariate and multivariate – for forecasting a set of Russian macroeconomic indicators. In the fourth chapter, the forecasting focuses on multivariate structural (DSGE) and non-structural BVAR models. The fifth chapter quantifies the effect of different types of oil market shocks on several Russian macroeconomic variables.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France