Università degli Studi di Cagliari

DOTTORATO DI RICERCA

ECONOMIA

Ciclo XXXVIII

TITOLO TESI

FORECASTING IN EASTERN EUROPEAN COUNTRIES

Settore/i scientifico disciplinari di afferenza

SECS-P/01 ECONOMIA POLITICA

Presentata da Dott.ssa Monika Junicke

Coordinatore Dottorato Prof. Romano Piras

Tutor Prof. Paolo Mattana

Esame finale anno accademico 2015 – 2016
Tesi discussa nella sessione d'esame marzo – aprile 2017

To My Beloved Family

Abstract

The thesis provides detailed empirical applications of two sets of forecasting methods, popular in the academic literature, using macroeconomic time series of three Eastern European countries (EECs): namely, the Czech republic, Hungary and Poland. The idea is to develop a natural extension to my previous studies, in particular those presented in Junicke (2017), where I applied Bayesian inference to produce an empirical estimation of a dynamic stochastic general equilibrium (DSGE) model for a small open economy. After discussing a survey of the literature on macroeconomics forecasting, which includes a discussion on the developments of methodology and accuracy measures, I first analyze the forecasts resulting from a model with theoretical grounds. Then, I turn to those resulting from a model with econometric foundations. My findings are twofold. First, it suggests that using different pure econometric models, allowing for parameters and covariance matrix to vary may improve the forecasting performance for EECs on average. Second, the DSGE models forecast better when trend inflation is explicitly taken into consideration.

Contents

Introduction

The contribution of this thesis is to provide detailed empirical applications of two sets of forecasting methods, popular in the academic literature, using macroeconomic time series of three Eastern European countries (EECs): namely, the Czech republic, Hungary and Poland. The idea is to develop a natural extension to my previous studies, in particular those presented in Junicke (2017), where I applied Bayesian inference to produce an empirical estimation of a dynamic stochastic general equilibrium (DSGE) model for a small open economy. After discussing a survey of the literature on macroeconomics forecasting, which includes a discussion on the developments of methodology and accuracy measures, I first analyze the forecasts resulting from a model with theoretical grounds. Then, I turn to those resulting from a model with econometric foundations. My findings are twofold. First, it suggests that using different pure econometric models, allowing for parameters and covariance matrix to vary may improve the forecasting performance for EECs on average. Second, the DSGE models forecast better when trend inflation is explicitly taken into consideration.

What do economist indicate with the term forecasting? According to a generally accepted definition, forecasting is a process yielding predictions about the future, based on an information set containing data about past and current events. Forecasting is thus permeated with uncertainty, characteristics that challenge the forecasting process itself. For its intrinsic nature, forecasting is nevertheless useful only if it is able to produce accurate predictions about future realizations of the variable of interest. A fundamental concern is therefore how effective is forecasting. In the aftermath of the financial crisis in 2008, it became clear that the existing forecasting techniques were performing poorly. Indeed, over the recent years, the performance of economic forecasting came to be a subject of many economic and political discussions. Reviewing the process through which we explain and predict comovements of macroeconomic variables turned into a major task for economists. Being used by most of the central banks, national governments as well as many international organizations such as the International Monetary Fund, World Bank and the OECD, in the last decade forecasting has therefore surged to an even more fundamental part of the macroeconomic research.

Unlike other branches of economic research, forecasting is not a field dominated by purely academic contributions. Being a central ingredient for enterprise success, formulating predictions about key macroeconomic variables is a process that many practitioners undertake. An interesting question is how the two approaches relate to one another. Reconciling professional forecasting to its academic counterpart is, however, an extremely complex task. The most important professional forecasting surveys, which include Blue Chips, Livingston Survey by Fed Philadelphia and World Economic Outlook by the International Monetary Fund (IMF) just to name some, publicly unfold neither the modeling assumptions they formulate, nor the information sets they use and the parameter estimates they produce. By and large for this reason, professional forecasting is mainly used by academics as a source of benchmark predictions against which to measure forecast accuracy.[1]

As mentioned above, uncertainty is the central issue of any forecasting process. "Forecasts [...] are unavoidably fraught with uncertainty due to imperfect understanding and observation of the mechanisms and forces determining economic outcomes" (Wieland and Wolters, 2013, p.241). Particularly in times when the volatility of macroeconomic variables is large, forecasting becomes very difficult. This is also the reason why only a few economists could predict the financial crisis that began in 2007. How good is forecasting today compared to then? Will we be able to predict the next economic crisis? How should we read market 'anomalies' such as the low crude-oil price, the slowing-down of the Chinese economy, the very low (and possibly negative) interest rates, the drop in the volumes exchanged

[1] For an example, see the survey on forecasting and policy making by Wieland and Wolters (2013).

on the international financial markets (notwithstanding the fact that these are overwhelmed with cheap money)? What will it be their effects on macroeconomic variables in the short and medium run?

The struggle in addressing these sort of questions is exacerbated when a limited amount of data is available. Under these circumstances, in fact, not only the preliminary process of estimating the model parameters is more troublesome, but also a fewer crisis episodes are documented. In fact, "macroeconometricians often face a shortage of observations necessary for providing precise answers. Some questions require high-dimensional empirical models. [...] Thus, sample information alone is often insufficient to enable sharp inference about model parameters and implications. Other questions do not necessarily require a very densely parameterized empirical model, but they do demand identification restrictions that are not self-evident and that are highly contested in the empirical literature. [...] Thus, documenting the uncertainty associated with empirical findings or predictions is of first-order importance for scientific reporting" (Del Negro and Schorfheide, 2011, p.294). These issues naturally pass on to forecasting. In order to reduce their extent, most of the existing work in the forecasting literature is based on countries, such as the U.S. or the U.K., for which rather large data sets are available.

Being based on a group of country for which reliable macroeconomic data have been made available only relatively recently, my thesis comprises instead empirical applications that rely on a severely limited number of observations.[2] While this is somehow a deliberate choice aimed at filling a gap in the forecasting literature and motivated by the common historical evolution of Czech, Hungarian and Polish economies, it nonetheless demands for an alternative solution to the uncertainly issues discussed above. In principle, to reduce the model's dimensionality "one could of course set many coefficients equal to zero or impose the condition that the same coefficient interacts with multiple regressors. Unfortunately, such *hard* restrictions rule out the existence of certain spillover effects, which might be undesirable. Conceptually more appealing is the use of *soft* restrictions, which can be easily incorporated through probability distributions for those coefficients that are "centered" at the desired restrictions but that have a small, yet nonzero, variance" (Del Negro and Schorfheide, 2011, p.296). Bayesian methodology fits naturally into this scenario: in Bayesian inference, instead of imposing a given value, a certain parameter is given a prior *distribution*, based on nonsample information, which is *updated* by sample information contained in the likelihood function to form a posterior distribution. For this reason, my choice is to resort to Bayesian methods for estimating the model and producing my forecasts. Although the Bayesian approach was practically absent from the literature on forecasting until a decade ago, as I illustrate in Chapter 1 it has nowadays surged to an active and vibrant area of macroeconomic research.

[2]The maximum range of available data for the three EECs span for about two decades.

Of course, everything comes at a cost. Consider the following example of identification problem. "Structural VARs can be parameterized in terms of reduced-form parameters, which enter the likelihood function, and an orthogonal matrix, which does not enter the likelihood function. Thus, [the orthogonal matrix] is not identifiable based on the sample information. In this case, the conditional distribution of [the orthogonal matrix] given the reduced-form parameters will not be updated, and its conditional posterior is identical to the conditional prior. Identification issues also arise in the context of DSGE models. In general, as long as the joint prior distribution of reduced-form and unidentifiable parameters is proper, meaning that the total probability mass is one, so is the joint posterior distribution. In this sense, the lack of identification poses no conceptual problem in a Bayesian framework. However, it does pose a challenge: it becomes more important to document which aspects of the prior distribution are not updated by the likelihood function and to recognize the extreme sensitivity of those aspects to the specification of the prior distribution" (Del Negro and Schorfheide, 2011, p.296). As a result, economists working with a Bayesian methodology must be very careful when performing a comparative analysis between prior and posterior distributions.

Opting for the use of Bayesian methodology poses no restriction in terms of modelling choice. Economists dealing with forecast analysis may choose from a large variety of econometric models, of which some may be univariate and others multivariate; some with fixed and others with time-varying parameters; some may be theory-based while others only make use of correlations in the data.[3] In the thesis, I consider a number of different models, which may be categorized by the intensity of theoretical background they incorporate. "The benefit of building empirical models on sound theoretical foundations is that the model delivers an internally consistent interpretation of the current state and future trajectories of the economy and enables a sound analysis of policy scenarios. The potential cost is that theory-implied cross-coefficient restrictions might lead to a deterioration in forecast performance" (Del Negro and Schorfheide, 2013, p.61). By examining predictions delivered by both types of model, the thesis offers an assessment of this trade-off when the empirical applications are based on relatively narrow datasets.

Regarding the forecasting model with theoretical background, Bayesian methodology has mostly been applied to Dynamic Stochastic General Equilibrium (DSGE) models. "DSGE models use modern macroeconomic theory to explain and predict comovements of aggregate time series over the business cycle. The term DSGE model encompasses a broad class of macroeconomic models that spans the standard neoclassical growth model discussed in King *et al.* (1988) as well as New Keynesian monetary models with numerous real and nominal frictions that are

[3]Methods of forecasting, alternative to econometric models and mostly used by practitioners, include economic base analysis, shift-share analysis, input-output models and the Grinold-Kroner model. Other, more specific, methods are land use forecasting, reference class forecasting, transportation planning, calculating demand forecast accuracy and consensus forecasts.

based on the work of Christiano *et al.* (2005) and Smets and Wouters (2003). A common feature of these models is that decision rules of economic agents are derived from assumptions about preferences, technologies, and the prevailing fiscal and monetary policy regime by solving intertemporal optimization problems. As a consequence, the DSGE model paradigm delivers empirical models with a strong degree of theoretical coherence that are attractive as a laboratory for policy experiments" (Del Negro and Schorfheide, 2013, p.58). Bayesian inference is used to investigate whether the predictions of the DSGE models match the statistical properties of the empirical data. This is crucial to identify the transmission channels for the exogenous shocks, and what is the behavior of governments and central banks in response to the latter. But it is also a powerful tool to forecast macroeconomic variables in a context where the model is also transparent with reference to the economic interpretation of the transmission channels.

Concerning the forecasting model with econometric background, Bayesian methodology has mainly been applied to Vector AutoRegressive (VAR) models. VARs "have become the workhorse model for macroeconomic forecasting. The initial use in economics was to a large degree motivated by Sims (1980) critique of the 'incredible restrictions' used by the large macroeconometric models developed in the 1970s and much effort was put into tools for policy analysis based on VAR models" (Karlsson, 2013, p.792). The thesis considers a number of different versions of econometric models, namely a linear vector autoregressive model (VAR) with fixed parameters, a threshold vector autoregressive model (TAR), a time varying parameters autoregressive model (TVP-AR) that serves as a benchmark model, and the time-varying parameters vector autoregressive model (TVP-VAR).

Drawing a comparison between DSGE and VAR models encompasses the evaluation of several aspects of the respective frameworks. The overall assessment typically vary substantially across economists, and comes to depend crucially on which aspects each places greater emphasis. For example, in the very same issue of the Handbook of Economic Forecasting (Volume 2), on the one hand Karlsson (2013, p.792) claims that the "role of the VAR model [as a tool for policy analysis] has to some degree been taken over by the current crop of DSGE models, a new generation of theory based models, which are —at times— ill at ease with the data. The role of the VAR model as the baseline, serious, model for economic forecasting is, however, unchallenged. The popularity stems in part from its relative simplicity, flexibility, and ability to fit the data but, of course, also from its success as a forecasting device." On the other hand, Del Negro and Schorfheide (2013, p.61) argue that the "case for DSGE model forecasting ultimately rests on the fact that these models provide a good package. Granted, there exist time series models that generate more accurate univariate forecasts of output growth and inflation, but these models might miss comovements between these two variables.

Bayesian VARs [...] tend to be good multivariate forecasting models but it is difficult to identify more than one or two structural shocks and to provide a narrative for the current and future state of the economy. Moreover, VARs typically do not have enough structure to generate predictions about anticipated changes in interest rates. [...] While the forecasting record of these models is strong, the policy experiments that could be carried out with these models are very limited. Finally, none of the aforementioned models would allow the user to measure the degree of distortion in the economy that ought to be corrected through monetary policy."

As a matter of fact, the last point made by Del Negro and Schorfheide (2013) is the reason why the thesis only looks into monetary policy issues in macroeconomic forecasting when using DSGE models. Introducing monetary policy shocks is fundamental to deliver forecasts conditional on specific interest rate patterns over time. "Central banks have long been concerned with estimating the transmission of changes in policy instruments such as the short-term nominal interest rate to target variables such as the rates of inflation and output growth. Their decisions are typically related to forecasts and adjusted frequently in response to new data, which induce forecast revisions" (Wieland and Wolters, 2013, p.241). Furthermore, considering external shocks such as monetary policy ones may help improving prediction reliability. "The accuracy of DSGE model forecasts is affected by how well the model captures low frequency trends in the data and by the extent to which important information about the current quarter (nowcast) is incorporated into the forecast. [...] [By introducing] shocks to the target-inflation rate, long-run productivity growth, as well as anticipated monetary policy shocks [...] [one may use] data on inflation, output growth, and interest rate expectations from the Blue Chip survey as observations on agents' expectations in the DSGE model and thereby incorporate the survey information into the DSGE model forecasts" (Del Negro and Schorfheide, 2013, p.62).

Naturally, forecast accuracy is of paramount importance for prediction reliability. For every model illustrated in the thesis, we examine the accuracy of our benchmark point forecasts, and whenever possible compare the precision of the predictions delivered by the models against a benchmark. This task is particularly important for DSGE forecasts. "Documenting the accuracy of DSGE model predictions is not just important from a forecasting perspective,but it can also be viewed as part of a general evaluation strategy for dynamic macro models. In fact, Smets and Wouters (2003, 2007) success of generating DSGE model forecasts that are competitive with forecasts from large Bayesian VARs convinced many researchers and policymakers that DSGE models are indeed empirically viable" (Del Negro and Schorfheide, 2013, p.62).

My thesis focuses on empirical applications of Bayesian forecasting methods, using both theoretically backed and purely econometric models, when only relatively short time series are available. The contribution of the thesis is twofold. First, it gives a complex overview over the quantitative forecasting methodologies, providing arguments against and in favour of each methodology. Second, it shows how different models, either purely econometric or with theoretical background, perform in terms of forecasting when data availability is severely limited.

The three Eastern European countries under consideration, because of their history, have in fact rather limited data sets. It is well known that the three countries developed in parallel. Their economies undertook a period of a central planning until the late 1980s, and engaged in an interesting transition path, including large structural changes in the 1990s, that lead the countries towards an open market economy. Since the econometrician can only rely on the data collected from the 1990s onwards, their data sets are obviously quite narrow. I exploit this feature to study the forecasting performance of Bayesian methodology under restricted data availability. Performing this task brings along the bonus feature of testing prediction reliability against data coming from a case study which proves unique not only from an economic perspective, but also from a geopolitical point of view.

The rest of the thesis is organized as follows.

In the first chapter, I give a detailed review of the literature on macroeconomic forecasting. The chapter draws substantially upon the excellent surveys by Del Negro and Schorfheide (2013), Karlsson (2013) and Wieland and Wolters

(2013). I introduce the leading forecasting models, separating my analysis of the contributions with theoretical grounds from those with econometric foundations. With regard to the first group, the focus is naturally on DSGE models of New Keynesian type. Concerning the second group, I consider a number of different versions of vector autoregressive models, including VAR with both fixed and time-varying parameters. I also survey some studies proposing combinations of the two approaches: the so-called pooled macroeconomic DSGE-VAR and DSGE-DFM models (DFM stands for dynamic factor model). I then review the contributions that developed the Bayesian inference and forecasting methodologies. I then explore the techniques to measure the models' forecasting accuracy, with reference to both point and density forecast. Finally, I briefly discuss the scant forecasting applications that use macroeconomic data of the Czech Republic, Hungary and/or Poland.

In the second chapter, I develop a simple DSGE model for a small open economy, and study its forecasting performance. The contribution may be seen as a natural extension of my work in Junicke (2017), where I apply Bayesian estimation techniques to investigate monetary policy in the three countries under consideration using a New Keynesian model. Building on this framework, I show that accounting for a steady state with non zero inflation significantly improves the forecasting performance of the model, regardless whether a recursive or rolling scheme is adopted.

In the third chapter, I study the performance of several purely econometric linear and non-linear models, namely a linear vector autoregressive model (VAR) with fixed parameters, a threshold vector autoregressive model (TAR), a time varying parameters autoregressive model (TVP-AR) that serves as a benchmark model, and the time-varying parameters vector autoregressive model (TVP-VAR). I show that allowing for parameters and covariance matrix to vary with time improves the forecasting performance, particularly on output growth. However, Threshold VAR allowing for the possibility of two different regimes may perform best in periods of recession.

Chapter 1

Literature review

Macroeconomic forecasting is the process of making predictions about variables that portray the economy and its performance at a high level of aggregation. Forecasting is one of the most interesting and important tasks for macroeconomists. At the same time, it is also one of the most challenging ones. The reason is that, for a forecasting model to be considered reliable, it has to deliver sound predictions precisely in those situations where economic outcomes are most uncertain. Is there a forecasting method or model that enable to predict large volatility in macroeconomic variables? In light of the Great Recession, the answer is evidently negative. Since the financial crisis that began in 2007, which the mainstream literature failed to predict, economists are looking for a forecasting method that will possibly predict the next one.

Uncertainty is a big issue in forecasting, and comes from two sources. The first source is the potentially unreliability of the data upon which the forecasting is based. Most of the data mining is accompanied by 'noise', which derive from sampling (surveys) or even the estimation processes (misspecification). The second source originates from deviations from the assumption that the forces that created the past event will be active in precisely the same way in the future. In fact, the prospective evolution of the macroeconomic variables is itself uncertain. They may be influenced by random elements in the future: *e.g.*, the development of a currently unpredictable new technology, policy changes, or simply some elements that currently look irrelevant but may turn out to be crucial in the future.

Thus, a model that is able to fit well the data sample may still fail in forecasting. The ensuing uncertainty about the model specification is well emphasized by Hirano and Wright (2017, p.617): "consider a pseudo-likelihood setting with a fixed number k of potential parameters to be estimated, each of which has a coefficient that is local to zero. The concept of model selection that we envision amounts to selecting a set of zero restrictions; in a regression setting, for example, this would indicate which predictors are excluded from the regression. Thus, there are up to 2^k possible models among which we can choose. Having chosen the model, we then have to estimate the parameters and use these for forecasting. Although some model will be best in terms of predictive accuracy, the local-to-zero nesting means that we can never consistently select that model."

In this perspective, Bayesian methodology lends a helping hand. Instead of setting coefficients equal to zero, this methodology allows for the use of 'soft' restrictions, incorporated in the probability distributions of coefficients that are 'centered' in zero but have perhaps small, yet non null, variance (Del Negro and Schorfheide, 2011). In the Bayesian framework, instead of imposing a given value, a certain parameter is given a prior distribution, based on nonsample information, which is updated by sample information contained in the likelihood function to form a posterior distribution.

How do we assess forecasting models during periods in which 'natural experiments' such as the Great Recession do not occur? The main criteria for judging the forecasting quality is based on the predictions' precision. There exist several methods to measure forecast accuracy. The most widely adopted ones are based on several variations of the root-mean-square error (RMSE), which measures the differences between values (sample and population values) predicted by a model and the values actually observed. It is often shown that weighted forecast, computed by combining models of different types, generally leads to more accurate forecasts than those delivered by individual models.

Indeed, a great many varieties of models populate the very broad forecasting literature. This has to do with the fact that virtually every economic organization, be it governmental or a large private institution such as a commercial or an investment bank, features a team or an advisory panel that uses macroeconomic models to forecast future development of key economic variables. In order to rationalize the structure of my literature survey, I split the contributions into three main categories.[1] Forecasting may either be based on models with a theoretical background, usually DSGE models featuring some version of the relationship known as the Phillips curve. Or it is performed using purely econometric models, such as autoregressive and vector autoregressive models. Or again, based on a combination of these two types of models.

This chapter reviews the most relevant contributions in on forecasting modelling. It also discusses the leading developments of Bayesian methodology, the literature on forecasting accuracy, and the works dealing with the three Eastern European countries under consideration throughout the thesis. Most of the discussion here focus on overviewing the works on forecasting models. I first discuss the studies based on frameworks with theoretical grounds. The focus is on dynamic stochastic general equilibrium models of New Keynesian type. I describe the main assumptions, paying particular attention to the supply side of the economy, the monetary policy and its targets. I also briefly illustrate the role of habit formation, and the importance of introducing a non zero trend inflation. I then turn to

[1] My literature review draws substantially upon the excellent surveys by Del Negro and Schorfheide (2013), Karlsson (2013) and Wieland and Wolters (2013).

survey models with econometric foundations, focusing on vector autoregressions. I outline the main features of models with both fixed parameter and time-varying parameters. Finally, I review the contributions considering a combination of the theoretical and econometric approach. The emphasis there is given to the so-called DSGE-VAR models and the DSGE-DFM models. I then turn to describe the works implementing the Bayesian methodology, which is used throughout the thesis. The description of Bayesian inference is followed by a survey on how to measure forecasting accuracy. Finally, I examine the rather limited number of contributions in the forecasting literature involving Eastern European countries.

By comparing different quantitative forecasting models, I summarize their relative advantages and disadvantages. The driver of my review is the question: what is the 'correct' model to produce forecasts? It is the million dollar question that researchers constantly try to answer in the literature. There is a large portfolio of models from which a forecaster can choose. The most frequently used in the current literature may be divided into two groups. The first group, comprising models with a rich economic structure such as the New Keynesian type, consists of DSGE models with market imperfections. These generally perform worse, but their advantage is the strong 'explanatory power', the capability of 'telling a story' that can be used to improve precision of analysis or to answer a policy question. The second group encompasses time-series econometric models, purely based on the observed relationships among the endogenous variables of the model. It is shown that, in general, they capture the relationships between the variables better than the first group of models, and they typically deliver better forecasts. Since both types of models display some relative advantage, it is not surprising that there exists a third group of models constructed as a mixture of the original two. In particular, the DSGE-VAR models take the form of a structural VAR, and allow the researcher to simultaneously estimate the parameters of the DSGE model and the VAR.

"Even if most private sector forecasters still favor traditional structural models over the New-Keynesian DSGE models with microeconomic foundations preferred by academia and central banks, the two types of models exhibit some similar reduced-form relationships such as price and wage-inflation Phillips curves and aggregate demand equations with a mixture of backward- and forward-looking components" (Wieland and Wolters, 2011, p.250). In fact, the New-Keynesian models are increasingly being used by central banks around the world as tools for macroeconomic forecasting and policy analysis. Examples of such models include the small open economy model developed by the Sveriges Riksbank (Adolfson *et al.*, 2007), the New Area-Wide Model developed at the European Central Bank (Coenen *et al.*, 2008; Christoffel *et al.*, 2011), and the Federal Reserve Board's New Estimated, Dynamic, Optimization-based model (Edge *et al.*, 2010). Forecasting with the DSGE models is nowadays based on Bayesian methods (see, for instance, An and Schorfheide, 2007; and, for a review, Del Negro and Schorfheide, 2010), particularly if the goal is to track and forecast macroeconomic time series. However, maximum likelihood estimation is also used (see, *e.g.*, Ireland, 2007).

Notwithstanding, vector autoregressive models may still be considered the workhorse for macroeconomic forecasting (Karlsson, 2013). "Many different multivariate time series models have been used in macroeconomics, but since the pioneering work of Sims (1980), Vector Autoregressive (VAR) models have been among the most popular" (Koop and Korobilis, 2009, p.268). There are several versions of VAR models, each resulting from different observational motivations. For instance, "in many applications, the assumption that the VAR coefficients were constant over time might be a poor one. [...] This led to an interest in models which allowed for time variation in the VAR coefficients and time-varying parameter VARs (TVP-VARs) arose" (Koop and Korobilis, 2009, p.268); or the "Great Moderation of the business cycle led to an increasing focus on appropriate modelling of the error covariance matrix in multivariate time series models and this led to the incorporation of multivariate stochastic volatility in many empirical papers" (Koop and Korobilis, 2009, pp.268-269).

Finally, it should be noted that a number of alternative methods of forecasting exist. These include economic base analysis, shift-share analysis, input-output model, the Grinold-Kroner model, land use forecasting, reference class forecasting, transportation planning, calculating demand forecast accuracy and consensus forecasts. Since these methods do not immediately relate to my work illustrated in the later chapters of the thesis, in what follows I do not discuss this professional-based branch of the literature.

1.1 Forecasting models with theoretical grounds

Authors advocating the use of theory-based models for forecasting have a rather passionate view about the reasons why predictions should result from that type of models. "Many macroeconomists have a strong preference for models with a high degree of theoretical coherence such as dynamic stochastic general equilibrium (DSGE) models. In these models, decision rules of economic agents are derived from assumptions about agents' preferences and production technologies and some fundamental principles such as intertemporal optimization, rational expectations, and competitive equilibrium. In practice, this means that the functional forms and parameters of equations that describe the behavior of economic agents are tightly restricted by optimality and equilibrium conditions. Thus, likelihood functions for empirical models with a strong degree of theoretical coherence tend to be more restrictive than likelihood functions associated with atheoretical models. A challenge arises if the data favor the atheoretical model and the atheoretical model generates more accurate forecasts, but a theoretically coherent model is required for the analysis of a particular economic policy" (Del Negro and Schorfheide, 2011, p.294)

The most popular DSGE models used to forecast the main macroeconomic variables, such as output and inflation, are those based on the Phillips curve. In the wake of Alban W. Phillips' (1958) postulated trade-off relationship between the rate of unemployment and the rate of inflation, many economists concluded that governments could lower unemployment by increasing inflation. It was assumed that, by appropriately setting the rate of inflation, monetary policy authorities could achieve the desired level of employment. However, in the 1970s, during the oil crisis, the idea of a fixed relationship between inflation and employment, which had already been questioned during the previous decade, was abandoned by most economists. The empirical observations showed no longer such a trade-off between inflation and unemployment, and stagflation took place instead.

Phillips' findings were suggestive of the possibility that monetary authorities should intervene to rectify the relationship between inflation and output. Neoclassical economists found this proposition somehow too 'static', as it did not account for the fact that individuals are able to form inflationary expectations based on past inflation. Their argument led to the traditional acceleration Phillips curve, characterized by backward looking components. After the famous critique by Robert

E. Lucas (1976), the theory was microfounded. Individuals based their choices on fully rational decisions. Building on this critique, Kydland and Prescott (1982) formulated a theory for microfounded business cycle models, known as the Real Business Cycle (RBC) theory. These authors theorized that productivity shocks are responsible for economic fluctuations, hence arguing that monetary policy has no real effects on the economy. Their model became the building block for the New Keynesian models (NKM), which represent the class of models that central banks currently most commonly rely upon.

Empirical tests of this new wave of models did not produce comforting results. Roberts (2001), Eichenbaum and Fisher (2003) and Dupuis (2004) showed that the standard new Keynesian Phillips curve (NKPC) with a simple forward looking component did not correspond well with empirical results, as it was unable to capture inflation persistence. Furthermore, Ball *et al.* (2005) pointed out that the NKPC lacks any source of inflation inertia and therefore it makes absurdly counterfactual predications about the effects of monetary policy. Indeed, models are modified for empirical estimation to contain (in most cases) a hybrid Phillips curve, though often not microfounded.

There exist several ways to achieve a hybrid Phillips curve through microfoundation. To incorporate observed inflation persistence, Galì and Gertler (1999) create a model that includes a hybrid Phillips curve with both forward and backward looking components, with the marginal cost as the driving variable. The firms are subject to constraints on the timing of their price changes, such that they can only change prices after a random interval of time has passed. It also allows firms to operate two types of pricing policies. The first has that firms attempt to maximize the discounted value of profits in the form of Calvo's (1983) contracts, while the second predicates that firms follow a simple rule of thumb, which updates prices in line with observed price changes and inflation. This model is based on micro-foundations, but it also includes an explanation for inflation persistence.

The most commonly used New Keynesian models belong to a group of dynamic general equilibrium models with uncertainty about the timing of price setting. These belong to the class of non-linear models with structural interpretable relationships between the variables. The cornerstone of such models is the fact that prices are staggered. Indeed, many empirical studies confirm the hypothesis that prices are not adjusted continuously. For most countries, the data suggests that prices change on average once a year.[2] Beside long term contracts, the main reason for staggered prices are menu costs, such as administrative costs, costs of re-labeling packages and costs for advertising these changes. Additionally, psychological effects of attractive prices (*e.g.*, a .99 tacked onto the end of the price) only allow for a step-wise change of prices.

[2]However, Gali and Gertler (1999), Kozicki and Tinsley (2002) and Sbordone (2002) show that, depending on the chosen method and assumptions, the average period may vary dramatically.

There are several options for introducing price stickiness into the model. The most commonly used mechanisms are those by Calvo (1983), Taylor (1983) and Rotemberg (1982). In particular, Calvo's model generates a standard purely forward looking NKPC. In this model, competition is imperfect in the good market, and prices are sticky. Although models based on Calvo's mechanism are frequently used to derive monetary policy rules, they are often criticized for their unrealistic assumptions. In this class of models, in fact, prices changes are independent over time. This feature contradicts a large body of empirical evidence showing that inflation is strongly autocorrelated. For instance, Angeloni *et al.* (2004) show that inflation inertia is significant in all European countries.[3]

The supply side, represented by the Phillips curve, is the most important and most controversial element of the New Keynesian DSGE models. The central bank seeks to stabilize the economy by setting goals and appropriate rules to achieve these goals. Given the central bank preferences and the actual fundamentals of the economy, summarized by inflation and output gap, this rule implies a target interest rate that may well differ from the one that would arise in the market (in the absence of central bank intervention). Should that be the case, the central bank would intervene in the financial market by purchasing (if the target interest rate

[3]The inflation persistence network (IPN), built by ECB and the 12 National Central Banks, focuses on measuring and comparing patterns of price setting and inflation persistence in the Euro Area. Angeloni *et al.* (2002) define inflation persistence as "the tendency of inflation to converge slowly (or sluggishly) towards its long-run value".

is lower than the free-market counterpart) or selling (otherwise) bonds. The final result of this process would ensure that the nominal interest rate in the financial market equals the target interest rate, and the actual state of the economy is characterized by a system of three equations (namely, the Euler equation, the Phillips curve and the interest rate targeting rule) in three variables (namely, the nominal interest rate, inflation and the output gap).

In the last forty years, the theory of monetary policy has increasingly attracted the attention of economists. Special attention has been paid to the trade-off between inflation and income variations. There are numerous examples of how monetary authorities has dealt with this issue in practice: during the 1980's the German Bundesbank targeted inflation by changing its money supply directly; in the 1990's, the Central Bank of New Zealand was the first bank to adopt an interest rate rule, then followed by many others authorities such as the Bank of England, the Sveriges Riksbank (the central bank of Sweden) and the European Central Bank (ECB).

In a SOE framework, the Taylor rule can be used in two different ways.[4] In fact, by moving the interest rate, the central bank can either target producer domestic inflation or consumer inflation. Galì and Monacelli (2005) and Sutherland (2002) point out that if the economy's non-stochastic steady state is in its optimum and no (or only very small) cost push distortions are present, the optimal monetary policy

[4] For further discussion, see Gali and Monacelli (2005) and Di Giorgio and Nisticò (2007).

is pure domestic inflation targeting. Strict producer-price targeting has a smoother effect on domestic variables without any distortion to the foreign economy.

Regarding the demand side of the economy, some researchers replace the traditional IS curve with an alternative one obtained from a utility maximizing framework with external habit formation, originally introduced by Fuhrer (2000). Since then, habit formation became an important extension to the baseline NKM. Habit formation is modelled by assuming that consumer's current utility is determined by current and lagged consumption in a non-time separable way. According to Fuhrer (2000), a model with habit formation fits the data better because it is able to deliver closer predictions to the empirically observed persistence in the consumption growth process. Including habit formation delivers hump-shaped impulse response functions, and prevents a jump response in real consumption expenditures, As a result, the model has a closer correspondence to reality and does not allow for inflation to be a jump variable.

Another important feature of NKM is the evolution of inflation over time. Although empirical analysis are not suggestive of a zero steady state inflation, it is common in the literature to assume zero trend inflation, which enables to log-linearize the models easily. However, there are several attempts in the literature to incorporate a non zero steady state inflation. First, Ascari and Ropele (2007) log-linearize a baseline NKM around a non zero inflation trend. They find that introducing a trend inflation changes the dynamic of key endogenous variables in the model, such as inflation and output growth. Another contribution to this subject, which also tackles the problem of zero lower bound on nominal interest rate, is Coibion *et al.* (2012). These authors study the optimal monetary policy rules and find that low positive inflation lowers the costs of the zero lower bound.

The literature using Bayesian techniques to estimate the model and to forecast the macroeconomic variables using DSGE models is large. The first important work in this field is Smets and Wouters (2003), who estimate structural parameters of a closed economy model using Euro Area data. This work has since been extended for small open economy models. Lubik and Schorfheide (2005) create a symmetric two-country model, and estimate it using U.S. and Euro Area data. Using a similar dataset, Rabanal and Tuesta (2010) estimate and compare models with complete and incomplete financial markets. In a later paper, Lubik and Schorfheide (2007) estimate how central banks in Australia, Canada, New Zealand and the UK respond to exchange rate changes, estimating composite structural parameters. Similarly, Adolfson *et al.* (2008) and Liu (2006) investigate similar questions while assuming incomplete pass-through, using data for Sweden and New Zealand, respectively. Justiniano and Preston (2010) identify the optimal policy rule within a generalized class of Taylor-type rule, which they estimate using data from Australia, Canada and New Zealand. They show that these rules do not respond to the nominal exchange rates. Del Negro and Schorfheide (2009) also study the effect of changes in the monetary policy rule, using data for Chile.

The central reference for DSGE based macroeconomic forecasting is Smets and Wouters (2007), who build on earlier contributions by Christiano *et al.* (2005) and Smets and Wouters (2003). "It is a medium-scale DSGE model, which augments the standard neoclassical stochastic growth model by nominal price and wage rigidities as well as habit formation in consumption and investment adjustment costs" (Del Negro and Schorfheide, 2013, p.63). Another key reference is the financial accelerator model developed by Bernanke *et al.* (1999), which complements the Smets-Wouters model by considering credit frictions (see also Christiano *et al.*, 2010).

The benchmark model can be enriched with a number of additional features. For instance, forecasts can be generated conditional on a certain level of interest rate by using a sequence of unanticipated monetary policy shocks, as in Leeper and Zha (2003) and Smets and Wouters (2005). "Leeper and Zha (2003) recommend to analyze the effect of monetary policy interventions with unanticipated shocks only if the interventions are modest. Here modest essentially means that in a larger model in which agents assign positive probability to occasional shifts in policy regimes, the intervention would not trigger the learning mechanism and lead the agent to believe that the policy regime has shifted" (Del Negro and Schorfheide, 2013, p.113).

This way of taking into account unanticipated monetary policy shocks has recently been extended. In fact, "the literature has considered the use of anticipated monetary policy shocks to generate forecasts conditional on an interest rate path that deviates from the model implied path, *e.g.*, Laseen and Svensson (2011), Blake (2012), and Milani and Treadwell (2012). This approach is appealing because several central banks have changed their communication strategy and started to announce interest rate paths" (Del Negro and Schorfheide, 2013, p.113).

Another possibility that has been examined in the literature is to depart from log-linearized DSGE models with Gaussian innovations: "starting with the work of Fernández-Villaverde and Rubio-Ramírez [2013], an increasing number of researchers have applied likelihood based estimation methods to DSGE models solved with non-linear techniques. To the extent that nonlinearities, *e.g.*, time-varying volatilities, kinks such as a zero-lower-bound constraint on the nominal interest rates, and business cycle asymmetries are empirically important features of macroeconomic time series, incorporating them into forecasting models is potentially beneficial. To date there have been few systematic studies of the forecasting performance of non-linearly solved DSGE models. The reason is that the estimation of non-linear DSGE models is much more time consuming than the estimation of linearized models, which makes it cumbersome to conduct recursive out-of-sample forecasting exercises" (Del Negro and Schorfheide, 2013, p.129).

One way to circumvent the computational issues arising with recursive forecasting of non-linear models is to consider linearized DSGE models with student-t (instead of Gaussian) shocks, as in Chib and Ramamurthy (2014) and Cúrdia *et al.* (2014). Another way is to replace the homoscedastic shocks in a linearized DSGE model by shocks that exhibit stochastic volatility, as in Justiniano and Primiceri (2008). "Their approach is promising for forecasting applications because, [with] the exception of the time-variation in the shock standard deviation, their solution concept still delivers a linear state-space model which is fairly easy to estimate. The inclusion of stochastic volatility makes the implied predictive distribution of the model more adaptive to changes in macroeconomic volatility" (Del Negro and Schorfheide, 2013, p.130).

Another consideration that surfaced in the literature is that "macroeconomic instabilities not only manifest themselves in volatilities but also in levels of macroeconomic time series. Correctly capturing such level instabilities in DSGE models is likely to lead better conditional mean forecasts. [...] Pichler (2008) conducts a systematic study of the role of nonlinearities for forecasting with a New Keynesian DSGE model with capital accumulation. His analysis consists of two parts. He shows that in simulations in which data are generated from a DSGE model that has been approximately solved with a second-order perturbation method, the forecasts from the non-linear version of the DSGE model are indeed more accurate than forecasts from a linearized DSGE model. On actual U.S. data, however, the result was reversed: the forecasts from the linearized model tended to be more precise than the forecast from the model approximated to second order" (Del Negro and Schorfheide, 2013, p.130).

Naturally, the objective is that non-linear models will be able to help forecast future crisis. Recent contributions, surveyed by Brunnermeier *et al.* (2013), suggests that "non-linear models with financial frictions potentially offer some promise, at least in terms of explaining the crisis without requiring very large shocks. So far, to our knowledge none of these models have been taken to the data using either calibration or econometric methods however, and we do know anything about their forecasting ability. Estimation of non-linear models with financial frictions is therefore an interesting area of future research" (Del Negro and Schorfheide, 2013, p.130).

The specification of DSGE models has dramatically improved in the last decade. Economists have been able to develop novel methods to incorporate real-time information, to relax cross-equation restrictions, and to combine DSGE models with other macroeconometric frameworks. "The progress is in part driven by the desire of central banks to incorporate modern macroeconomic equilibrium theory into their decision-making process. In this regard, the recent crisis with the emergence of non-conventional monetary policies and interest rates near the zero-lower bound has supplied new challenges for DSGE model-based forecasting that need to be tackled in future research. [...] [Also,] financial markets data can contain valuable information for forecasting. A further challenge for DSGE model forecasters will be to exploit these data. In order to do that, it is first necessary to build models that have a shot at explaining these data, by itself a tall order. Non-linear models,which can generate time-varying risk premia, hold some promise in this dimension" (Del Negro and Schorfheide, 2013, pp.133-134).

The amount of research evaluating the accuracy of point forecasts from DSGE models is rather large. Many of the studies consider versions of Smets and Wouters (2003, 2007) models, which can this be considered as the benchmark for this literature. Several contributions in the DSGE model forecasting literature compare their forecasts to those produces with other models. Edge and Gürkaynak (2010) offer a comparison between univariate forecasts from the benchmark model forecasts obtained from the Federal Reserve, the Blue Chip survey, and a Bayesian vector autoregression (VAR). The exercise is run with real-time data, and the comparison is based on RMSEs. The authors find that the DSGE model performance is competitive in terms of accuracy with those obtained from the alternative prediction methods. Contributions comparing DSGE model and professional forecasts can also be found in Wieland and Wolters (2011, 2013).

The evidence from the Euro Area data is similar. Adolfson *et al.* (2007) investigate how an open economy DSGE model performs in terms of forecasts during the decade 1994-2004. Accuracy measurements are based on RMSEs, log determinant of the forecast-error covariance matrix, predictive scores, and the coverage frequency of interval forecasts (see Section 1.5 for details). The authors find that, overall, the DSGE model performance is in line with more those of flexible time series models such as VARs. As mentioned above, Christoffel *et al.* (2011) study the forecasting performance of the New Area Wide Model (NAWM), the DSGE model used by the European Central Bank. The authors assess the model's univariate forecast accuracy using RMSEs, and the multivariate forecast accuracy using log determinant of the forecast-error covariance matrix. They find that the DSGE model performance is similar to those obtained with other forecasting models such as VARs of various sizes.

Herbst and Schorfheide (2012) simulate a DSGE model, then generate recursive forecasts on the simulated trajectories. This way, they obtain the predictive distribution for RMSEs implied by the model. "The authors find that, for a small-scale DSGE model, the actual RMSEs of output and inflation forecasts are within the bands of the predictive distribution. The actual interest rate RMSEs, on the other hand, exceed the predictive bands, indicating a deficiency in the law of motion of the interest rate. For the Smets and Wouters (2007) model, the inflation and interest rate RMSEs fall within the bands of the predictive distribution, but the realized output growth RMSE is smaller than the RMSE predicted by the model. A possible explanation is that some of the estimated shock processes are overly persistent because they need to absorb violations of the balanced growth path restrictions of the DSGE model. This would lead to excess volatility in the simulated output paths" (Del Negro and Schorfheide, 2013, p.90).

Summarizing, the empirical evidence in the literature indicate that DSGE model forecasts are in line with those produced by standard autoregressive or vector autoregressive models. However, they can be dominated by more complex models. At any rate, DSGE models have some advantages relative to reduced form models since they offer an economic explanation of the predictions. Furthermore, DSGE models also generate a framework that is suitable to undertake policy analysis. In forecasting, this feature may be important since DSGE models can be used to make predictions based on alternative patterns of the policy instruments.

1.2 Forecasting models with econometric foundations

If authors opting for theory-based forecasting models could be deemed to have quite a strong view about the reasons behind their choice, economists favouring econometric models might sound almost dismissive. "Vector autoregressions (VARs) have become the workhorse model for macroeconomic forecasting. The initial use in economics was to a large degree motived by Sims (1980) critique of the 'incredible restrictions' used by the large macroeconometric models developed in the 1970s and much effort was put into tools for policy analysis based on VAR models. This role of the VAR model has to some degree been taken over by the current crop of DSGE models, a new generation of theory based models, which are —at times— ill at ease with the data. The role of the VAR model as the baseline, serious, model for economic forecasting is, however, unchallenged. The popularity stems in part from its relative simplicity, flexibility, and ability to fit the data but, of course, also from its success as a forecasting device" (Karlsson, 2013, p.792). Naturally, not every economist agree upon this view. Various criticism about Bayesian VARs arise regarding difficulties emerging when attempting to identify more than one or two structural shocks, or to provide a narrative for the current and future state of the economy, or again to generate predictions about anticipated changes in interest rates.

VARs are econometric models that capture the linear interdependencies among multiple time series. They represent the natural generalization of the univariate autoregressive model (AR model), obtained by considering more than one evolving variable. Historically, VARs might be considered as the evolution of simultaneous equation models (SEMs). There, variables are distinguished between endogenous (dependent variables) and exogenous (independent variables). The data are left to 'dictate' the relationships between endogenous and exogenous variables. The VAR approach generalizes this framework by considering each variable in a symmetric fashion. All variables involved are allowed to be both independent (current value) and dependent (lagged values) variables. A VAR model has therefore a very simple structure: each variable is assumed to be dependent on its own lagged values, the lagged values of the other model variables, and an error term. VAR modeling does not require much knowledge about the forces influencing a given variable. The only prior knowledge required is a list of variables which can be assumed to affect each other intertemporally. A slightly more demanding version of vector autoregressive models is the structural VAR, which adds cross-variables contemporaneous relationships to the lagged ones that characterize simple VARs. Structural VARs may, however, be given a simple VAR representation by performing straightforward algebraic transformation. The resulting framework is known as reduced-form VAR.

Perhaps the greater weakness of VAR models is the risk of overfitting the data, which entails large uncertainty about the future realizations of the variables predicted by the model. With the term 'overfitting' economists refer to the potential tendency of a statistical model to misinterpret random error or noise as part of an underlying relationship between variables. VARs might suffer this occurrence when the model specification is excessively complex, that is to say, too many parameters are included in the analysis. In the attempt to prevent this issue, macroeconometricians have often resorted to the Minnesota prior. Introduced by Litterman (1979), the Minnesota prior is a set of prior beliefs that push the parameter values towards a stylized representation of macroeconomic data. By reducing parameter uncertainty, the introduction of this prior is bound to improve forecast accuracy. "The Minnesota prior captures widely held beliefs about the long-run properties of the data, properties that are not readily apparent in the short samples typically used for estimation" (Karlsson, 2013, p.793).

Another possibility to constrain parameter values in order to reduce uncertainty is the Bayesian methodology. According to this methodology, prior beliefs on the probability distribution of parameters values are updated using information gathered from data. Bayesian method has also the advantage that the macro-econometrician may control the tightness of the adherence to prior beliefs that the parameters values must fulfil. A possibility to merge the long-run beliefs with the information gathered from of the data is represented by the so-called vector error-correction models (VECM). (This type of models are not discussed in the thesis, yet they represent a promising tool for future research on the issues discussed in the next chapters.) The implementation of VECMs typically involves the estimation of unrestricted VARs, 'corrected' to use possible cointegration between variables as a source of information for their long-run relationship. Together with, or instead of, the observed cointegration, an economist could also include beliefs like the Minnesota prior. Also in this respect, Bayesian methodology plays a central role in the technical implementation of the model. "Bayes theorem then provides the optimal way of combining these two sources of information leading to sharper inference and more precise forecasts. The development of efficient numerical techniques for evaluating posterior distributions is also a contributing factor to the attractiveness of Bayesian methods. It is now possible to tackle more complex problems under realistic assumptions when we no longer are limited to problem formulations that lead to analytical solutions" (Karlsson, 2013, p.793).

A relatively recent development in the VAR literature has been to assume away that the regression parameters are constant, a feature characterizing the vector autoregressive models as well as a range of DSGE models with fixed parameters. Relaxing this assumption allows for time varying parameters and stochastic volatility in Bayesian VAR models. The advantage is that the macroeconometrician may be able to pinpoint structural breaks in the relationships between variables, which could in turn be the result of changes in the economy's technology or in the policies implemented by policy makers. The disadvantage is that it may exacerbate the overfitting issue of the model. "There are encouraging studies that indicate that both time-varying parameters and stochastic volatility can improve the forecast performance but both can also lead to a dramatic increase in the number of parameters in a model. There is consequently a greater risk of overfitting the data" (Karlsson, 2013, p.794).

The literature on VAR forecasting is extremely large, and virtually span from the beginning of the 1980s, in the wake of Sims' (1980) seminar work on the approach, to date. Instead of getting lost in the myriad of different existing applications, it is convenient to mention that the most influential early contributions include Litterman (1986) and Sims (1993). These authors' primary objective was finding ways to control the number of parameters in large VAR models. The issue is a very serious one. Small (two- to three-variable) VARs are useful as a benchmark, yet they are often unstable and thus poor predictors of the future (Stock and Watson, 1996). "State-of-the-art VAR forecasting systems contain more than three variables and allow for time-varying parameters to capture important drifts in coefficients (Sims, 1993). However, adding variables to the VAR creates complications, because the number of VAR parameters increases as the square of the number of variables: a nine-variable, four-lag VAR has 333 unknown coefficients (including the intercepts). Unfortunately, macroeconomic time series data cannot provide reliable estimates of all these coefficients without further restrictions" (Stock and Watson, 2001, pp.110-111). The idea they pursued is to impose a common structure on the coefficients using Bayesian methods (respectively, on a six-variable and on a nine-variable system). Building on these efforts, later forecasting systems have displayed solid real-time track records (McNees, 1990; Zarnowitz and Braun, 1993).

More recently, the approach that produced perhaps the most substantial advancement is the one introducing time-varying parameters (TVP). A number of early studies considered parameter variations in VAR models (for the early stages, see, *e.g.*, Sims, 1993; Doan *et al.*, 1984, who conducted their analysis using the Kalman filter; and Highfield, 1987, who relaxes this assumption and uses the normal-Wishart conjugate prior; and also Chib,1998, who accommodates structural breaks by including a Markov switching mechanism with a fixed number of regimes; more recently, see Pesaran *et al.*, 2006, and Koop and Potter, 2007, who contemplate an evolving number of regimes; Benati, 2010, and Mumtaz and Sunder-Plassmann, 2013, who use time-varying VARs to capture the time-varying dynamics of U.K. macroeconomic and financial time series). The boost in the popularity of TVP-VARs is, however, probably due to the introduction of Bayesian methodology, pioneered by Cogley and Sargent (2002, 2005) and Primiceri (2005) who, although not primarily concerned with forecasting, provide the foundations for Bayesian inference in these models.

A number of influential applications of Bayesian VARs (BVARs) are worth mentioning. Canova (2007) forecasts the inflation rate of the G7 countries using a range of models, including BVARs and Bayesian panel VAR. He concludes that the later performs best, also exploiting the largest available information sets; a BVAR improves on a VAR estimated with OLS; and time-varying parameters improve the forecasts for univariate models, but not for the BVARs. Clark and McCracken (2010) use a real-time data set and forecasts the U.S. inflation, interest rate, and output to produce forecasts using a wide range of trivariate VAR models differing on the approach adopted and on whether they allow for structural change. BVARs and TVP-BVARs are included in the set. These authors find that a BVAR with detrended inflation does best and, while not directly comparable, considerably better than a TVP-BVAR where inflation has not been detrended. D'Agostino *et al.* (2013) show that a VAR with time-varying parameters and stochastic volatility performs well in forecasting U.S. macroeconomic data.

Another assumption that has been questioned in the literature is the constancy of the error variance. This issue is particularly relevant since, with the advent of the 'Great Moderation', macroeconomic variables began to exhibit considerably lower variability from the mid-1980s. It is not an easy task to empirically distinguish a model with constant parameters and time-varying variances from a model with time varying parameters and constant error variance. It is therefore customary to prudentially allow for both. Both Cogley and Sargent (2005) and Primiceri (2005) consider time varying parameter VAR models where also stochastic volatility is introduced.

More recently, Clark (2011) produces point and density forecasts using real-time data on U.S. output growth, unemployment rate, inflation, and federal funds rate. Under the conjecture that the Great Moderation might be over, the author investigates the effects of the ensuing changing data variability using a model that combines a BVAR as in Villani (2009) with stochastic volatility *à la* Cogley and Sargent (2005). The model forecast performance is compared to those of AR models, with and without stochastic volatility, and of BVARs with Minnesota priors. All BVARs are estimated on both a recursively updated and rolling data window. The author finds that BVARs generally perform worse than the benchmark AR model without stochastic volatilities in the short run, yet their performance improve on the in the medium and long run. The results are also suggestive of a positive role of stochastic volatility in further improving forecasting performance. This appears to be a more general result than it may initially seem. In fact, "the stochastic volatility models do considerably better at the shorter lead times while the differences are quite small for the longer lead times. Similarly the steady-state BVAR outperforms the standard BVAR at short lead times and the steady-state BVAR with stochastic volatility outperforms the steady-state BVAR at shorter lead times" (Karlsson, 2013, p.858).

Another example, already mentioned above, of a forecasting exercise that investigates on the importance of considering time-varying parameters and stochastic volatility is D'Agostino *et al.* (2013). Forecasts are on U.S. macroeconomic variables such as unemployment rate, inflation, and short-term interest rate, and are based on several models: a standard AR, a SV-AR and a TVPSV-AR; a SV-VAR and TVPSV-VAR using Primiceri's (2005) specification. Using Bayesian inference and prior beliefs based on the Minnesota prior, the authors show that the TVPSV-VAR performs best both in terms of point forecasts and density forecasts. Also, the SV-AR and the SV-VAR models improve on their constant variance counterparts. These results allow the authors to conclude that there is a role for both time-varying parameters and time-varying error variances when forecasting the most important macroeconomic variables.

A factor that has an important impact on forecasting performance is the choice of the prior distribution. Using reduced form VARs, Litterman (1986) and McNees (1986) assess the forecast accuracy resulting from adopting Minnesota prior of true out-of-sample forecasts compared to commercial forecasts based on large scale macroeconometric models. Their finding is mixed. On the one hand, BVAR forecasts do better for the real variables (real GNP, investments, and unemployment); on the other, they performed poorly for inflation and the T-bill rate.

The model size also influence forecasting performance. Giannone *et al.* (2015)

forecast the U.S. GDP, GDP deflator, and federal funds rate, using three different BVARs implemented using hierarchical prior and VARs estimated with OLS, a random walk with drift and a dynamic factor model. In terms of mean square error, they find that BVARs improve with the size of the model. This is in contrast to the OLS estimated VARs. The largest BVAR produces better one-step-ahead forecasts than the factor model for the GDP deflator and the federal funds rate, and better four-step ahead forecasts for the GDP deflator.

Regarding structural VARs, Österholm (2008) uses a structural BVAR for Sweden and offers a limited forecast evaluation. In the forecast evaluation, a steady-state version of the SVAR and a naive random walk are also considered. The author finds that the steady-state SVAR produces the best forecasts for Swedish inflation, with the forecast performance of the structural BVAR improving on that of the random walk. For GDP growth, the steady-state SVAR is again best, while the structural BVAR performance is the worst. The random walk provides the best forecasts for the inflation rate followed by the steady-state SVAR; once again, the poorest performance is the one by the structural BVAR.

Villani (2001) forecasts the Swedish inflation rate with several versions of VECMs. The author considers several theory based cointegrating relations, which are all rejected by the data. Nonetheless, comparing forecasts with both theory based (PPP, stationary domestic, and foreign interest rates) and unrestricted cointegrating vectors with stationary Minnesota type prior beliefs on the short run dynamics, he finds that Bayesian VECMs perform better than maximum likelihood estimated VECMs and an ARIMA model.

Already Sims (1993) reports on the enhancements made to the original Litterman forecasting model when allows for conditional heteroscedasticity and non-normal errors in addition to time varying regression parameters. The result of these modifications are an improved forecasting performance for nominal variables and comparable or slightly better forecasts for real variables compared to the original Litterman model. As discussed above, Canova (2007) forecasts the inflation rate of the G7 countries using a range of models, including a BVAR with Minnesota style prior beliefs, a Bayesian TVP-VAR and a Bayesian TVP-AR with mean reverting state equation. The author finds that, overall, the model with the largest information set and the most general specification, the Bayesian panel VAR, does best. Comparing models with similar information sets, the BVAR improves on the VAR estimated with OLS, and time-varying parameters improve the forecasts for univariate models but not for the BVARs. Clark and McCracken (2010) use a real-time data set and forecasts the U.S. variables using a wide range of trivariate VARs, BVARs with Minnesota type priors and TVP-BVARs with random walk state equations. The authors report that BVAR with detrended inflation performs best and,while not directly comparable, considerably better than a TVP-BVAR where inflation has not been detrended.

Korobilis (2008) applies a stochastic search variable selection (SSVS), initially proposed by George *et al.* (2008), in a forecasting VAR model. Testing the model on U.S. macroeconomic data, the author finds that the SSVS model predictions improve on the forecasts from OLS estimated VARs. Jochmann *et al.* (2010) proposes to implement the SSVS procedure onto VAR models with Markov switching, in order to allow for structural breaks. Using U.S. data on unemployment, interest rate, and inflation, the authors find that the restriction search leads to a better performance than a BVAR with a 'loose' prior or with a Minnesota prior. Korobilis (2013) introduces SSVS in a larger set of multi-variate time series models. Based on U.K. data unemployment, interest rate, and inflation, the author finds that the restriction search improves forecast performance when the prior is informative, the model is large of highly parameterized.

Using a lager set of U.S. data, Carriero *et al.* (2011) compare the performance of Bayesian procedures (VAR with normal-Wishart prior, the reduced-rank VAR and the reduced-rank posterior) with several alternatives (a reduced-rank VAR estimated with maximum likelihood, multi-variate boosting, factor models and univariate autoregressions). They find that the reduced-rank posterior and the Bayesian reduced-rank VAR procedures perform best, both in terms of overall forecasting and when for specific variables of interest.

1.3 Forecasting with combinations of models

As I discussed in the first section of the chapter, the benefit of building empirical models on sound theoretical foundations is that the model delivers an internally consistent interpretation of the current and future states of the economy and allows for a solid scrutiny of potential policy scenarios. Limitations, however, apply. While the evidence collected in several studies indicates that DSGE models may be taken seriously from a forecasting perspective, it should be kept in mind that the number of studies is still quite limited and that the forecast samples considered do not cover events, such as a deep recession, that are particularly difficult to foresee (Christoffel *et al.*, 2011).

In this respect, as I discussed in the second section of the chapter, econometrically founded models seem to offer a better performance. Nevertheless, they suffer of identification issues and provide a less qualified framework for investigating possible policy scenarios. For this reason, a number of contributions attempt to combine the two lines of modelling to exploit the respective advantages, and minimize the disadvantages. In what follows, I briefly illustrate the two leading combinations surfacing the literature, namely: vector autoregressions with restriction generated by dynamic general equilibrium models (DSGE-VAR); and dynamic factor models (DSGE-DFM).

Combination of DSGE models with VARs

The first contribution is due to Del Negro and Schorfheide (2004), who propose to combine a general equilibrium model with a vector autoregression (VAR) to obtain a framework delivering forecasts that both perform well and are usable for policy analysis. In particular, their approach uses prior information coming from a DSGE model in the estimation of a VAR. Based on a hierarchical Bayes model, the framework is known as DSGE-VAR, takes the form of a structural VAR and allows the researcher to simultaneously estimate the parameters of the DSGE model and the VAR. A hyperparameter determines the scale of the prior covariance matrix. If the prior covariance matrix is zero, then the DSGE model restrictions are dogmatically imposed on the VAR.

Quite a few contributions have followed the seminal article by Del Negro and Schorfheide (2004). Amisano and Geweke (2013) consider a pool of macroeconomic models that incorporates DSGE models and VARs. A combined forecast is generated from a convex combination of the predictive densities associated with the models included in the pool. The authors find that the DSGE model receives a substantial weight in the model 'portfolio', and that the forecasting performance of the 'portfolio' is substantially better than those of any individual model. Waggoner and Zha (2012) build on this approach by allowing for time-varying weights that follow a regime-switching process. The authors' analysis suggest that situations can be identified in which a substantial weight on the DSGE model is useful for macroeconomic forecasting, and others in which the 'portfolio' forecasts perform better when the predominant weight is assigned to the VAR. A similar approach can be followed in terms of density forecasts. Gerard and Nimark (2008) combine density forecasts of a DSGE model, a FAVAR model and a Bayesian VAR. Given the bad performance of individual models' density forecasts, it comes at no surprise that combined density forecasts overestimate uncertainty as well.

A Bayesian approach to VAR that uses DSGE model restrictions is to construct a micro-founded prior about VAR parameters and thus may improve VAR estimates by incorporating extra information. Alternatively, this method can be viewed as a way to improve the empirical properties of the DSGE model by relaxing tight cross-equation restrictions that might be at odds with real data. The idea of the approach is to simulate data from the DSGE model, append simulated to actual data and estimate a VAR on the extended sample. The optimal proportion (which can be estimated) of simulated to actual data measures the weight on DSGE restrictions. In this respect, comparing density forecasts of DSGE models with the actual distribution of observations, Wolters (2011) shows that the models overestimate uncertainty around point forecasts. Also, Kolasa *et al.* (2012) show that the DSGE model actually outperforms their DSGE-VAR specification.

More generally, however, DSGE-VAR models deliver more precise predictions than the underlying DSGE models: see, for instance, Del Negro *et al.* (2007) for a variant of the Smets and Wouters (2007) model, and Warne *et al.* (2014) for the New Area Wide Model (NAWM) of the European Central Banks. Ghent (2009) produces forecasts from DSGE-VARs that have been developed merging information from different DSGE models. She finds that the forecasting performance is alike across model specifications. This is suggestive of the fact that it may not be the particular economic structure, but rather general implications about data persistence that yield better forecasting performance such as those attained by DSGE-VARs.

The Bayesian VARs that serve as a benchmark in the DSGE model forecast literature typically use a version of the Minnesota prior. This prior lacks some of the later optimizing empirical findings described in Sims and Zha (1998) and, more recently, in Del Negro and Schorfheide (2011). Del Negro and Schorfheide (2004) use Bayesian VARs with prior distributions centered at the DSGE model restrictions. By letting the variance of the prior vary, these authors modify the weight placed on the DSGE model restrictions. They find that the resulting DSGE-VAR performs significantly better than the underlying DSGE model in terms of forecasts.

Combination of DSGE models with DMFs

Another possible combination is the one between DSGE and dynamic factor models (DFMs). Dynamic factor models were originally proposed by Geweke (1977) as a time-series extension of factor models previously developed for cross-sectional data. The founding idea of a dynamic factor model is that a few latent dynamic factors drive the comovements of a high-dimensional vector of time-series variables, which is also affected by a vector of mean-zero idiosyncratic disturbances. In early influential work, Sargent and Sims (1977) showed that two dynamic factors could explain a large fraction of the variance of important U.S. quarterly macroeconomic variables, including output, employment, and prices. Later, Sargent (1989) showed that the DFM can be interpreted as relating multiple indicators to a latent low-dimensional model of the economy.

The objective of combining DFM and DSGE models is to build a bridge between theoretically based frameworks and a large cross section of macroeconomic variables, rather than a small group of observables. This has two potential advantages. First, the large set of macroeconomic variables might provide sharper inference about the current state of the economy. Second, this framework allows the modeler to assess the effect of structural shocks, *e.g.*,monetary policy shocks, on variables that are not explicitly modeled in the DSGE model. The resulting empirical specification is called DSGE-DFM. It is essentially a DFM in which the latent factors are equated with the state variables of a DSGE model and follow the DSGE model-implied law of motion (Stock and Watson, 2011).

The DSGE-DFM is due to Boivin and Giannoni (2006), who link the dynamic factor evolution to a log-linearized DSGE model. Accordingly, the factors correspond to latent economic variables such as inflation and the output gap. The econometrician does not observe these latent variables, he only produce a number of measures of them, for instance, rates of inflation derived from different price indices. These observables constitute the set of model variables, and the factors are pinpointed by exclusion restrictions in the factor loadings (for example, the multiple observed measures of inflation depend directly on the latent inflation factor but not on the other factors).

Kryshko (2011) builds on the work of Boivin and Giannoni (2006), and show that the space spanned by the factors of a DSGE-DFM is very similar to the space spanned by factors extracted from an unrestricted DFM. The author integrated a medium-scale DSGE model into a dynamic factor model for a large cross section of macroeconomic indicators, thereby linking 'non-core' variables to a DSGE model. He jointly estimated the DSGE model parameters and the factor loadings. As a result, his factor estimates have clear economic interpretation. The joint estimation, conceptually very appealing, has a limit in its computational complexity, which makes it impractical for several types of forecasting exercises.

A recent, simpler approach is due to Schorfheide *et al.* (2010), who use a DSGE-DFM to generate DSGE model based forecasts for variables that do not explicitly appear in the DSGE model. The predictions for the 'non-core variables' are obtained by applying their measurement equations to DSGE model-generated forecasts of the state variables. The authors use a medium-scale New Keynesian DSGE model, and produce forecasts for, *e.g.*, inflation and unemployment rate. They find that, while their approach does not lead to a substantial reduction in forecast errors, predictions are competitive with those of the benchmark autoregressive model and are produced in a time frame that makes them suitable also for real time analysis.

All in all, the results in the empirical literature on forecast combination show that combining multiple forecasts increases the forecasting accuracy. Unless one can identify a single model with superior forecasting performance, forecast combinations are useful for diversification reasons as one does not have to rely on the predictions of a single forecasting model.

1.4 Bayesian Inference

DSGE models and VARs are successful tools in several branches of macroeconomic research, particularly to perform structural analysis and forecasting. These models allow to develop a general representation of the relationships between the time series of interest and make predictions about their future evolution. Using Bayesian methodology benefits estimation of and forecasting with both DSGE models and VARs, though for diametrically opposite reasons. As far as DSGE models are concerned, Bayes theorem allow to relax the tight restrictions resulting from the economic structure of the framework. With regard to VARs, characterized by a high number of parameters on which no restriction is imposed, the Bayesian approach provides a solution to overcome these problems by imposing restrictions through the prior distribution.

Bayesian methodology allow the macroeconometrician to decide how tight the restrictions to the model should be, regardless whether these restrictions originate from a theoretical model or are imposed *ad hoc* to the econometric model, through the choice of the parameters that control the strength of the prior belief on the posterior distribution. Another advantage of this method is the fact that, via the prior distribution, the macroeconometrician is able to exploit additional information in the estimation and forecasting of the model, rather than limiting the parameter estimation only to the information contained in the considered data sample.

There exist several sources of information that are approximately independent of the data that enter the likelihood function, and therefore could be used for the 'refinement' of the prior distribution specification (Del Negro and Schorfheide, 2011). First, information from macroeconomic variables different from those explicitly considered in the dataset. Second, microeconomic observations that may complement the information about the macroeconomic variables included in the analysis. Third, macroeconomic data on the same variables considered in the dataset relative to other periods than the one object of analysis.

The application of Bayesian inference to macroeconometric studies is relatively recent. This is mostly due to its algebraic complexity; the computation of the posterior distribution is based on integral calculus, which rarely yield analytical solutions. In order to avoid incurring in these algebraic issues, most of the results have been historically based on specific classes of probabilities for which the posterior distribution has the same shape as the prior. Thanks to the greater availability of computer resources since the early nineties, this restrictive approach has been gradually abandoned, as the improving computational power made it possible to solve the integrals numerically.

This possibility has also stimulated the application of the Bayesian statistical numerical methods developed in other contexts, such as those based on simulations (Monte Carlo method, sampling algorithms Gibbs and Metropolis-Hastings), as well as the development of new methods in the field of Bayesian statistics itself (for example, the popular methods based on Markov Chain Monte Carlo or MCMC). The Gibbs sampling algorithm relies on the availability of conditional distributions to be operational. In many cases (of practical relevance) conditional distributions are not available in closed form. An important example of such a situation is the estimation of DSGE models, where the conditional distribution of different parameter blocks is unavailable. In such cases an algorithm more general than the Gibbs sampler is required to approximate the posterior distribution. The Metropolis-Hastings algorithm offers such an alternative (Blake and Mumtaz, 2012).

In order to compute the mean and variance of the simulated posterior distribution, it is often convenient to resort to the use of a Kalman filter. The Kalman filter is an efficient recursive filter that evaluates the state of a dynamic system from a series of measurements subject to disturbance. Because of its inherent characteristics, the Kalman filter represent a solid tool for noise reduction on zero mean Gaussian systems.

The choice of the prior distribution on the structural parameters is an important part of the estimation, and should be kept in line with the existing literature whenever possible. The leading contribution for a closed economy is Smets and Wouters (2003). The priors are usually tight for parameters that are not in the main interest of the estimation, and as such can be used to impose economic beliefs. Conversely, priors for the core parameters of the estimation have significantly higher confidence intervals, and may be restricted, only if necessary, according to their economic interpretation.

To avoid the problem of stochastic singularity, the number of observed variables has to be equal to the number of uncorrelated shocks in the model. This may be problematic in a large scale model, since in this case it is required to add shocks that are not necessarily structured. Therefore, it is common to add some disturbances interpreted as a measurement error, although there may be some difficulties in the identification of such shocks within an economic context.[5]

The literature focuses on the estimates of the structural parameters, like those in the NKPC and in the Euler equation, and especially on the size of the backward looking components. Most of the estimates suggest a hybrid NKPC with a backward looking component. The value of lagged inflation should be around 0.3 to 0.5, in line with other empirical findings such as Galì and Gertler (1999) and Galì *et al.* (2001). Benigno and López-Salido (2006) test the Galí-Gertler model on five countries in Euro Area, showing that Germany firms behave rationally and, therefore, inflation is strongly forward looking. By contrast, in France, Spain and Italy, the opposite is true. In these countries, firms are characterized by backward looking price setting behavior, strongly linking their prices to past conditions. These asymmetries are important as they complicate the determination of unique monetary policy within the Euro Area. However, some authors, *e.g.,* Levine *et*

[5]For more details, see Lubik and Schorfheide (2007).

al. (2012), show that including habit formation improves the performance of the model more than a backward looking component of the Phillips curve. The estimates for habit formation are usually high: the literature refers to a value around 0.8 (see, *e.g.*, Liu, 2006).

However, the focus of the literature lies mainly on the monetary policy analysis, such as the estimation of the parameters of monetary policy rules. Also, the Bayesian estimation enables to point out what kind of theoretical monetary policy rule fits the data best, and allows to determine to what extent different shocks drive the dynamics of the endogenous variables. Using the posterior of marginal data density, which can be interpreted as the maximum log-likelihood value, the model fit can be tested. Smets and Wouters (2003) study the transmission of monetary policy shocks and find out that there is a difference in the behavior between a temporary and a persistent monetary policy shock. The behavior of the temporary shock is in line with the literature, as it increases the interest rate leading to a decline in output and price level. The persistent shock, however, rules out the liquidity effect, hence the nominal interest rate drops immediately because the inflationary expectations are lowered.

Also, the change in monetary policy is permanent and therefore credible, so the expectations have time to adjust and the effect on output is relative small. Lubik and Schorfheide (2005) estimate parameters for closed and open economy given

U.S. and Euro Area data to study the dynamic of the exchange rate. Similarly, Adolfson *et al.* (2008) estimate the dynamic development on the Euro Area data. These studies show that the exchange rate dynamics depends mostly on PPP and monetary policy shock. Additionally, Lubik and Schorfheide (2007) study the monetary policy rules given the data of various open economies, such as Australia, New Zealand, the UK and Canada, to explore to what extent the central banks include exchange rate movements into consideration. They conclude that except for Canada, the central banks do not respond to exchange rate variations.

Generally, in the literature a growing number of contributions implement the Bayesian methodology for the estimation and forecasting of macroeconomic models. For instance, Bańbura *et al.* (2010) show how the Bayesian approach is suitable for large VARs, and find that the predictive accuracy of even a small VAR can be improved with the inclusion of new information resulting from the use of Bayesian methodology. Cogley *et al.* (2005) use a Bayesian VAR to evaluate and designate the monetary policies adopted in the U.K., including local approximations of inflation trend and volatility. The authors illustrate the mechanism through which different monetary policies influence inflation under predetermined regimes.

1.5 Forecasting accuracy

Accuracy is the predominant criteria used to assess forecast quality. There exist a number of tools to measure forecast accuracy, but the most widely adopted one in macroeconometric studies is the root mean square error (RMSE). RMSE is a measure of the differences between values (sample and population values) predicted by a model and the values actually observed. A variation of this indicator is root mean square deviation (RMSD), which reports the sample standard deviation of the differences between predicted and observed values. The individual differences are called residuals when the calculations are performed over the data sample that was used for estimation, and are called prediction errors when computed out-of-sample. The RMSE serves to aggregate the magnitudes of the errors in predictions for various times into a single measure of predictive power. RMSE is a good measure of accuracy, but only to compare forecasting errors of different models for a particular variable and not between variables.

An alternative accuracy measure for Bayesian model is the log predictive density score, which in particular represents a measurement of the forecast accuracy of the density. This measure is based on the posterior predictive distribution. This distribution is unobserved conditional on observed data and is derived by computing maximum likelihood estimate of the parameters given the observed data. These parameters are plugged into the distribution function of the new observations. In order to compare the models, the econometrician first computes the log predictive density score of the benchmark model, then the one for the alternative model. The second model performs better than the first at some forecast horizon simply if the score is positive.

Another measure relies on the marginal likelihood. In statistics, marginal likelihood is a likelihood function in which some parameter variables have been marginalized. In the context of Bayesian statistics, it can be used to compare the performance of Bayesian models. The marginalized variables are parameters for a particular type of model, and the remaining variable is the identity of the model itself. In this case, the marginalized likelihood is the probability of the data given the model type, not assuming any particular model parameters.

A forecaster willing to evaluate the goodness of interval predictions (Chatfield, 1993) may instead opt for the coverage frequency of interval forecasts. In a nutshell, this measure 'counts' the occurrences in which the actual observations lies within the forecast interval. Naturally, a model that is capable of producing narrow intervals in tranquil times and wide in volatile times is bound to outperform a model delivering 'unconditional' intervals, since the occurrences of observations outside the interval forecast would be spread out over the sample and not come in clusters.

Forecasting of the DSGE models can in fact be made on point, volatility, interval or density. Usually, a DSGE model is solved in a log-linearized form and it is assumed that volatilities are constant, although there are many examples in the literature where the disturbances are time varying. In any case, as it is pointed out by, *e.g.*, Diebold *et al.* (2016), density forecast generated from a DSGE model is typically too wide and, as such, it is preferable to concentrate on point forecast. Point forecast is traditionally based on the average tendency of a variable.

Bayesian inference offers the advantage that the it allows for closed form solution of the density function. Generally, the higher the density function, the better the fit of the model to the data. Consider a variable of interest and a forecast for it made at some point in time. The forecast accuracy is measured by the expected loss, a function that reflects the type of forecast being performed: a point, density or interval forecast. It is common to use a quadratic function, or a function of the deviations in absolute terms. The forecast is produced out of sample, *i.e.*, the data sample is divided into two sub-samples, where the first one contains the first R observations. These data are used to estimate the parameters, which are in turn used to produce an h step ahead forecasting. This forecast is then compared to the observations in the second sub-sample.

A density forecast consists of computing forecast draws by iterating over the solution matrices for different parameter values drawn from the posterior distribution. At each iteration, a vector of shocks is drawn from a mean zero normal distribution with the variance itself being a draw from the posterior distribution. The forecast density is given by the ordered forecast draws. The econometrician can then compute the mean of the forecast density to get the point forecast. Using only one model to produce forecasts is equivalent to imposing a subjective prior of the forecaster implying that the specific model is the best representation of the unknown true data generating process. Combining forecasts from several DSGE models and unconstrained Bayesian VARs improves the robustness of the predictions, and the mean forecast tend to become as precise as the best nonstructural forecast.

A number of methods can be used to combine forecasts from the set of models: likelihood based weights, relative performance weights based on past RMSEs, a least squares estimator of weights, and non-parametric combination schemes. Overall, it appears that model combination methods that give weight to several models improves forecasting performance. Weighted forecasts have a higher accuracy than forecasts from individual models. Combined forecasts based on simple weighting schemes that give significant weight to several models are superior to likelihood based weighting schemes that turn out to identify a single model rather than giving weight to several models. While point forecasts are interesting, economists are concerned about the uncertainty surrounding the predictions. Therefore, density forecasts are somehow preferred for the DSGE models. These take into account parameter uncertainty and uncertainty about economic shocks in the future (Wolters, 2011).

Assuming a symmetric loss function, the accuracy of point forecasts can be easily compared by computing RMSEs. Evaluating density forecasts is less straightforward. The true density is never observed. It is based on the relationship between the data generating process and the sequence of density forecasts via probability integral transforms of the observed data with respect to the density forecasts. In this respect, a useful tool is the probability integral transform (PIT), which is the cumulative density function corresponding to the sequence of a set of density forecasts evaluated at the corresponding observed data points. The PIT is the probability implied by the density forecast that a realized data point would be equal or less than what it is actually observed. If the sequence of density forecasts is an accurate description of actual uncertainty, the sequence of PITs should be distributed uniformly between zero and one. The empirical studies in the literature suggest, however, that PITs tend to overestimate actual uncertainty.

1.6 Eastern European Countries data and forecasting

Since the fall of the Berlin Wall, the Eastern European countries (EECs) have undergone significant structural changes. This was the result of the transition from a centrally planned economy like the socialist system to a free market economy. The foremost changes introduced by market liberalization were the removal of all price controls, and the gradual privatization of small and large business, which until then were state-owned. These occurrences were followed by the development of the private business sector, which in turn resulted in the restructuring of entire industries (*e.g.*, heavy industry).

The Polish economy is the biggest among the EECs under consideration. Between 1989 and 2007, the Polish economy grew by 177%, faster than other countries in Europe (Eastern, Central and Western alike). This was the result of a 'shock therapy' programme, initiated by Leszek Balcerowicz in the early 1990s, which enabled the country to reach its pre-1989 GDP levels, which it achieved by 1995 as the first post-communist country. Poland exhibits a strong domestic market, low private debt, flexible currency, and is not dependent on a single export sector. As a result, it was the only European economy to have avoided the late-2000s recession. The Polish currency is the złoty (PLN) . It underwent a redenomination on January 1, 1995, as a result of which the old currency by a new one: the redenomination rate was 10,000 old Polish złoty to 1 new Polish złoty. Since then, the currency has been relatively stable, with an exchange rate fluctuating between 3 and 4 złoty for a U.S. dollar.

In November 1989, Czechoslovakia returned to a liberal democracy through the peaceful 'Velvet Revolution'. When Slovak national aspirations strengthened, the country peacefully split into the independent Czech Republic and Slovakia. Both countries went through economic reforms and privatization, with the intention of creating a market economy. This process was largely successful, and in 2006 the Czech Republic was recognized by the World Bank as a 'developed country'. The Czech Republic is high-income economy with a per capita GDP rate that is 87% of the European Union average. It also has the lowest unemployment rate in the European Union. Monetary policy is conducted by the Czech National Bank, whose independence is guaranteed by the Constitution. The official currency is the Czech koruna. The Czech crown became fully convertible for most business purposes in late 1995. Following the Czech economic troubles, which culminated in a currency crisis in May 1997, the formerly pegged currency was forced into a floating system. In November 2013, the Czech National Bank started to intervene to weaken the exchange rate of Czech koruna through a monetary stimulus in order to stop the currency from excessive strengthening and to fight against deflation.

On April 6th, 2017, the CNB stated the return to conventional monetary policy.

Hungary's transition from communism to democracy and capitalism was also peaceful and prompted by economic stagnation and domestic political pressure. With the removal of state subsidies and rapid privatization in 1991, Hungary entered a severe economic recession. The government's austerity measures proved unpopular, yet the government privatization program ended on schedule in 1998. The program seems to be paying off: Hungary continues to be one of the leading nations for attracting foreign direct investment in Central and Eastern Europe. Hungary maintains its own currency, the Hungarian forint (HUF), although the economy fulfills the Maastricht criteria with the exception of public debt (which is anyway significantly below the E.U. average). In 1995 the forint (HUF), became convertible for all current account transactions. In 1996, convertibility was extended to almost all capital account transactions as well. Since 1995, Hungary pegged the forint against a basket of currencies (in which the U.S. dollar has a 30% share). The central rate against the basket was devalued at a preannounced rate, originally set at 0.8% per month. The forint is now an entirely free-floating currency. The Hungarian National Bank—founded in 1924, after the dissolution of Austro-Hungarian Empire—is currently focusing on price stability with an inflation target of 3%.

It should be noted that, under the terms of their Treaty of Accession to the

European Union, all new Member States "shall participate in the Economic and Monetary Union from the date of accession as a Member State with a derogation", which means that all countries are eventually obliged to replace their currency with the euro. This fact has entailed that, in governing the transition processes, policy makers had to shape their intervention with the objective of gradually becoming part of the European Union single market. The consequences of these considerations are reflected in the analysis of the EEC economies, regardless whether the analysis is undertaken by academics or practitioners, and are particularly important when the behavior of central banks is studied. This feature will be taken into account throughout the thesis.

Not surprisingly, the international academic literature lacks of contributions on forecasting of EECs macroeconomic variables. This is most likely due to the relatively shortage of data discussed above. However, the central banks of those countries, along with no doubt several other governmental and private institutions, do produce forecasts based on models analogous to those discussed in this chapter. For example, the central bank of Hungary publishes macroeconomic forecasts since 2001, and has since then used formal macroeconomic models to produce their predictions. At the early stages, the forecasting procedure was delivered by traditional econometric models. In the wake of the growing implementation of DSGE models in the academic world as well as among practitioners, the central bank of Hungary followed suit, and in 2011 introduced their first DSGE type model (Békési et al., 2016). The Czech National Bank forecasts are based on a structural model of a small open economy, which is based on observed Czech data as well as additional assumptions that include the functioning of foreign economies, fiscal policy and administered price outlooks, the short-term forecast of the exchange rate and inflation. The National Bank of Poland produces their predictions, under the assumption of constant NBP interest rate, using a structural macroeconometric forecasting model designed to describe the Polish economy, developed in 2008.

1.7 Concluding remarks

There is a large literature on both the theoretical and empirical background of macroeconomic forecasting. The chapter has offered a detailed review of this literature. I have illustrated the most important developments about forecasting models, splitting my analysis of the contributions with theoretical grounds from those with econometric foundations. With regard to the first group, the focus has naturally been on dynamic general equilibrium models of New Keynesian type. The review has shown that there has been a significant improvement of these models in terms of forecasting performance, though the prediction accuracy of alternative models is still generally superior. Nevertheless, DSGE models are considered in the literature as an essential tool for addressing forecasts involving economic policy analysis.

Concerning the second group, I have considered a number of different versions of vector autoregressive models, including VAR with both fixed and time-varying parameters. The study also described structural VARs and the relevant reduced-forms. Naturally, both OLS based and Bayesian VARs have been analyzed. My survey has shown that econometric models are still the best performing in the forecasting arena. Many issues, including the key identification problem, generating criticism towards these models have been dealt with and significantly reduced. Nonetheless, limits appear to still apply when considering complex models, particularly those designed to perform policy analysis. Overall, these models stand as the leader in the macroeconomic forecasting business.

I have also surveyed some studies proposing combinations of the two approaches: the so-called pooled macroeconomic DSGE-VAR and DSGE-DFM models (DFM stands for dynamic factor model). I have then reviewed the contributions that developed the Bayesian inference and forecasting methodologies. I have explored the techniques to measure the models' forecasting accuracy, with reference to both point and density forecast. Finally, I have briefly discussed the scant forecasting applications that use macroeconomic data of the Czech Republic, Hungary and/or Poland.

Chapter 2

Forecasting with DSGE models

This chapter examines the forecasting performance of two DSGE models for a Small Open Economy (SOE). In developing the models, I choose to include the macroeconomic variables that are most relevant to the Eastern European countries (EECs) under consideration, namely the Czech Republic, Hungary and Poland. The main theoretical model builds on the New Keynesian literature with non zero trend inflation. The structure of the model closely relates to those adopted in Junicke (2017), Galì and Monacelli (2005), Rabanal and Tuesta (2010) and De Paoli (2009). The forecasting performance of the model are then compared to that of a second New Keynesian model (NKM) with zero steady state inflation. I take this second model as a benchmark since the assumption of zero steady state inflation is one of the most common in the mainstream literature on SOE modelling.

In developing *both* models —with and without zero steady state inflation— I generalize the typical SOE framework by considering a number of additional assumptions. First, I introduce incomplete pass-through of the exchange rate to the local currency prices, as in Monacelli (2003), as well as a home bias in the representation of consumer preferences. These features lead to deviations from purchasing power parity (PPP), which is empirically more relevant than assuming that PPP holds. Furthermore, the intratemporal elasticity of substitution between domestic and foreign goods differs from unity, allowing the SOE central bank to manipulate the terms of trade (in turn related to the relative domestic price). The reason for introducing these variations to the models is twofold. On the one hand, Devereux and Engel (2003) show that optimal monetary policy, in case of less than perfect (incomplete) pass-through, should involve some consideration of exchange rate volatility. On the other hand, although it is typically assumed in the literature that the elasticity of substitution between domestic and foreign goods is one (as in, *e.g.*, Corsetti and Pesenti, 2001; Devereux and Engel, 2003; and Obstfeld and Rogoff, 2002), empirical estimations suggest larger elasticities. Using this result, Sutherland (2006) argues that the central bank should add the exchange rate as a monetary policy target.

The supply side is characterized by a hybrid New Keynesian Phillips Curve, which is derived using a rule of thumb following Galì and Gertler (1999). A similar

Phillips Curve specification is also used by Benigno and López-Salido (2006), who analyze the effect of asymmetric supply shocks across countries within a monetary union. Additionally, I follow Ascari and Ropele (2007) by log-linearizing the SOE Phillips curve around a non-zero steady state, and show that this assumption improves the fit of the model significantly. The monetary policy is specified as a simple targeting rules of Taylor type containing speed of inflation and the exchange rate.

I discuss my forecasting methodology and results in Section 2.2. There is a large literature using Bayesian techniques to forecast macroeconomic variables using DSGE models. The first important work in this field is Smets and Wouters (2003), who estimate structural parameters of a closed economy model using Euro Area data. This work has since been extended to SOE modelling. Lubik and Schorfheide (2005) create a two symmetric country model and estimate it using U.S. and Euro Area data. Using a similar dataset, Rabanal and Tuesta (2010) estimate and compare models with complete and incomplete financial markets. In a more recent paper, Lubik and Schorfheide (2007) estimate how central banks in Australia, Canada, New Zealand and the U.K. respond to exchange rate variations, estimating composite structural parameters. Similarly, Adolfson *et al.* (2008) and Liu (2006) investigate similar questions while assuming incomplete pass-through, using data for Sweden and New Zealand, respectively. Justiniano and Preston

(2010) identify the optimal policy rule within a generalized class of Taylor-type rule, which they estimate using data from Australia, Canada and New Zealand. They show that these rules do not respond to the nominal exchange rates. Del Negro and Schorfheide (2009) also study the effect of changes in the monetary policy rule, on Chilean data, using a DSGE-VAR model.

As always, the forecasting performance of a DSGE model is lowered by the fact that the model is just a simplified projection of the actual economy. Although this paper is in line with the existing literature, it nonetheless allows for some novel features to try and improve adherence of the model to the real world. In particular, the model features a Calvo mechanism within a Phillips curve with trend inflation, which is derived analytically. The fact that trend inflation is added to the model, departing from the seminal work by Smets and Wouters (2003), leads to an additional parameter, the elasticity of substitution between the goods, to be estimated. There is a large literature in the early 2000's discussing the role of this parameter, particularly because it determines the size of the mark-up charged by the monopolistically competitive firms.

In developing the structure of the model, I make every effort to appropriately balance two opposing factors. On the one hand, the model should be complex enough to approximate the actual structure of the estimated economy. On the other, too complicated models require large number of parameters, increasing the estimation errors and therefore lowering the forecast performance. For this reason, I refrain from using a larger-scale model such as the one developed by Smets and Wouters (2003), who include capital and investment as well as wage rigidities. My objective here is to focus on the essential dynamics of the model. This also has the advantage that the results are most adherent, and thus easily comparable, to those discussed in Chapter 3. Therefore, I carefully construct my model to appropriately fit the data, yet avoiding to over-complicate the framework.

The rest of the chapter is organized as follows. In Section 2.1, I specify the model assuming two asymmetric countries differing in size. After describing the demand and supply side of the model in details, I specify monetary policy as a nominal interest rate rule for each country, and log-linearize the model around its non-zero inflation steady state. In Section 2.2, I describe the estimation and forecasting methodology, the dataset, and the choice of prior. I also present the estimation results and some robustness test. Section 2.3 concludes.

2.1 The model

I specify the model for a small open economy (SOE), which interacts with a large economy. Section 2.1.1 describes in detail the household preferences, its optimization problem as well as total and aggregate demand for both the domestic and the foreign country. Section 2.1.2 illustrates the supply side of the model, and how the model is log-linearized around its steady state. The monetary policy rules in simple form are described in more detail in Section 2.1.3.

2.1.1 Demand side

I consider two countries: the home country, H, represents the SOE; the foreign economy, F, is sufficiently large to receive no influence by the SOE.[1] Consumption C is a Dixit-Stiglitz aggregator of home and foreign goods. In country H, at time t, consumption is formally given by

$$C_t = \left[(1-\lambda)^{\frac{1}{\theta}} \left(C_{H,t} \right)^{\frac{\theta-1}{\theta}} + \lambda^{\frac{1}{\theta}} \left(C_{F,t} \right)^{\frac{\theta-1}{\theta}} \right]^{\frac{\theta}{\theta-1}} \tag{2.1}$$

with θ denoting the intratemporal elasticity of substitution between domestic and foreign goods. The parameter $\lambda \in [0,1)$ is the degree of openness of country H. For the foreign economy, the quantity of imports from the SOE are assumed to be sufficiently marginal to assume $C_t^* = C_{F,t}^*$.[2] The domestic price index equation can be written as

$$P_t = \left[(1-\lambda) \left(P_{H,t} \right)^{1-\theta} + \lambda \left(P_{F,t} \right)^{1-\theta} \right]^{\frac{1}{1-\theta}} \tag{2.2}$$

[1] A detailed appendix about all the equations of the model is available upon request from the author.

[2] Starred variables are associated with the foreign economy. Generally, they are expressed in foreign currency. However, this rule does not apply to consumption, which is expressed in real terms: in this case, it is only used to distinguish between consumption at home and abroad.

For the foreign economy there is no dispersion between producer and consumer price index, hence $P_t^* = P_{F,t}^*$.

The domestic representative agent preferences are represented by a CRRA utility function as in Galì and Monacelli (2005). From the first order condition of the maximization problem of the domestic representative household, I derive the Euler equation for the domestic economy

$$E_t\left(\Pi_{t+1}\left[\frac{C_t}{C_{t+1}}\right]^{-\sigma}\frac{\varepsilon_t}{\varepsilon_{t+1}}\right) = \beta R_t \tag{2.3}$$

where $\Pi_{t+1} = P_{t+1}/P_t$ is a domestic CPI inflation, $\beta \in (0,1]$ is the subjective discount factor and R_t is the gross return on a riskless one year nominal bond. Following Steinbach *et al.* (2009), the expression $\varepsilon_{t+1}/\varepsilon_t$ can be interpreted as a risk premium on asset holding, *i.e.*, the wedge between the actual return on assets and the interest rate set by the central bank.

I assume that labour is immobile across countries. The domestic households labour supply is

$$\tilde{W}_t = \frac{N_t^\eta}{C_t^{-\sigma}}$$

where $\tilde{W}_t$ is the real domestic wage.

Assuming that the foreign household faces the same maximization problem, the Euler equation and the labour supply for a foreign economy are expressed analogously.

Because of the strong empirical evidence that the law of one price (LOP) does not hold, I assume incomplete pass-through. The LOP gap is therefore defined as

$$\Psi_t = S_t \frac{P_{F,t}^*}{P_{F,t}} \tag{2.4}$$

where the nominal exchange rate S_t denotes the price of the foreign currency in terms of the domestic currency.[3] Additionally, given the different degrees of home bias in consumption between the two countries, PPP does not hold, and the CPI differs across countries. Hence, the real exchange rate can be expressed as the price of foreign goods in term of domestic goods, that is

$$RS_t = \frac{S_t P_t^*}{P_t} \tag{2.5}$$

The relationship between domestic and CPI inflation is

$$\frac{\Pi_{H,t}}{\Pi_t} = \frac{\tilde{P}_{H,t}}{\tilde{P}_{H,t-1}} \tag{2.6}$$

where $\tilde{P}_{H,t} = P_{H,t}/P_t$ is the producer relative price. The relationship between imported and CPI inflation can be expressed as current relative to past import

[3] Note, however, that from the point of view of domestic producers the low of one price holds, because the relevant price is the one "at the dock".

prices, expressed in domestic currency,

$$\frac{\Pi_{F,t}}{\Pi_t} = \frac{\tilde{P}_{F,t}}{\tilde{P}_{F,t-1}} \tag{2.7}$$

with $\tilde{P}_{F,t} = P_{F,t}/P_t$.

Total demand for a generic domestic good i

$$Y_t(i) = \left(\frac{P_{H,t}(i)}{P_{H,t}}\right)^{-\varepsilon} \left(\tilde{P}_{H,t}\right)^{-\theta} C_t \left[1 - \lambda + \lambda RS_t^{\theta - \frac{1}{\sigma}}\right] \tag{2.8}$$

depends on the openness of the domestic economy λ, the dispersion between producer i price and the domestic producer price index $P_{H,t}(i)/P_{H,t}$, the dispersion between domestic producer and consumer price indexes $\tilde{P}_{H,t}$, and the real exchange rate RS_t. Note that a real depreciation of the exchange rate leads to an increase in production of good i.

The aggregate demand for domestic goods is

$$Y_t = \left(\tilde{P}_{H,t}\right)^{-\theta} C_t \left[1 - \lambda + \lambda RS_t^{\theta - \frac{1}{\sigma}}\right] \tag{2.9}$$

and the aggregate demand for goods produced in the large foreign economy is $Y_t^* = C_t^*$.

In my model, I ignore transaction costs and assume that financial markets are such that consumers from either country have access to both domestic and foreign bonds. The market price of a domestic riskless bond equals the expected discounted nominal return of the bond, formally $1/R_t = E_t\left[Q_{t,t+1}\right]$. Similarly for a foreign bond expressed in domestic currency, it holds that $S_t/\left(R_t^*\right) = E_t\left[S_{t+1}Q_{t,t+1}\right]$. With no possibility of arbitrage, the expected returns of these two bonds must be equal. Therefore, the uncovered interest parity equation can be written as the expected change in the real exchange rate and the ratio between domestic and foreign real interest rate

$$\frac{R_t}{R_t^*}E_t\left[\frac{\Pi_{t+1}^*}{\Pi_{t+1}}\right] = E_t\left[\frac{RS_{t+1}}{RS_t}\right] \tag{2.10}$$

Under the assumption of complete securities markets, consumption risk is perfectly shared and the stochastic discount factor, expressed in the same currency, is equal across countries. Assuming a zero steady state net demand for foreign assets and an *ex-ante* identical environment, I obtain the optimal risk sharing condition under complete financial markets

$$RS_t = \left(\frac{C_t^*}{C_t}\right)^{-\sigma}\frac{\varepsilon_t^*}{\varepsilon_t} \tag{2.11}$$

Therefore, deviations from power purchasing parity (PPP) imply different consumption levels across the two countries, caused by the changes in the real exchange rate. The difference between the foreign and the domestic preference shock

$(\epsilon_t^*/\epsilon_t)$ captures the deviations from optimal risk sharing.

2.1.2 Supply side

The supply side of the domestic economy consists of two parts. There are producers and import retailers, both setting prices in the manner described by Calvo (1983) and Galì and Gertler (1999). Each producer (resp., retailer) belongs to one of two types of firms. A measure $1-\omega$ set the price optimally, and are labelled f. A measure ω set the price according to a rule-of-thumb, and are labelled b. Firms may face two different situations: i) either they are allowed to set their price, with probability $1-\alpha$; ii) or they are not allowed to do so, with probability α.

The optimal choice of $(1-\alpha)(1-\omega)$ firms that can set their price at time t is

$$\tilde{P}_t^f(i) = \frac{J_t}{H_t} \tag{2.12}$$

where $\tilde{P}_t^f(i) = P_t^f(i)/P_t$ is the relative forward looking price of the domestic firm i. The numerator is

$$J_t = \mu V_t \left(C_t^{-\sigma} Y_t \widetilde{MC}_t \tilde{P}_{H,t} \right)^{\eta+1} + \alpha\beta E_t \left[\left(\Pi_{H,t+1} \right)^{\varepsilon} J_{t+1} \right]$$

where $\mu = \varepsilon/(\varepsilon-1)$ is the domestic mark-up, V_t is the mark-up shock and $\widetilde{MC}_t =$

$MC_t/P_{H,t}$ is the real marginal cost. The denominator of (2.12) is

$$H_t = C_t^{-\sigma} Y_t + \alpha\beta E_t \left[\left(\Pi_{H,t+1}\right)^{\varepsilon} \left(\Pi_{t+1}\right)^{-1} H_{t+1} \right]$$

The remaining $(1-\alpha)\,\omega$ domestic firms set prices at time t according to the rule of thumb

$$P_t^b = \Pi_{H,t-1} X_{t-1} \tag{2.13}$$

where X_{t-1} denotes an index of the prices set at date $t-1$, generically expressed by

$$X_t \equiv \left[(1-\omega)\, P_t^{f(1-\varepsilon)} + \omega P_t^{b(1-\varepsilon)} \right]^{\frac{1}{1-\varepsilon}} \tag{2.14}$$

The aggregate producer price level then follows the law of motion

$$P_{H,t} = \left[(1-\alpha)\, X_t^{1-\varepsilon} + \alpha \left(P_{H,t-1}\right)^{1-\varepsilon} \right]^{\frac{1}{1-\varepsilon}} \tag{2.15}$$

The set of equations (2.12)-(2.15) constitute the hybrid New Keynesian Phillips curve (NKPC), which characterizes the producer side of country H. The NKPC for importing retailers is derived analogously, following from the fact that the importers face monopolistic competition. For the large country F, the set of equations leading to the NKPC is derived similarly, though without the dispersion between PPI and CPI.

Steady State and Log-linearized Form of the Model

The structural equations can be written in the log-linearized form around their steady state. This is assumed to be a perfect-foresight state for both economies, with zero income growth and stable technology. I assume that in the steady state all prices change at the same rate, and the price of the imports increases at the same rate as the price of the domestically produced goods. I can normalize the price indices by imposing $P_H = P_F$. Therefore, it follows that the consumer and producer price index are equal, formally, $P = P_H$. Inflation as well as the relative prices do not change, and it holds that $\Pi_H = \Pi_F = \Pi = \Pi^*$.

The log-linearized equations characterizing the non-policy part are as follows. The domestic Euler equation (2.3) can be rewritten in terms of deviations from the steady state as

$$\hat{c}_t = E_t\left[\hat{c}_{t+1}\right] - \frac{1}{\sigma}\left(\hat{\imath}_t - E_t\left[\hat{\pi}_{t+1}\right] + E_t\left[\Delta\epsilon_{t+1}\right]\right) \tag{2.16}$$

where I have used the approximation $log\left(R_t\right) \approx \hat{\imath}_t$. The term $\Delta\epsilon_{t+1} = \log\varepsilon_{t+1} - \log\varepsilon_t$ is the first difference of the structural preference shock. The linearization of the uncovered interest parity (2.10) delivers the relationship between real interest rate $\hat{\imath}_t$ and real exchange rate $\widehat{rs}_t$

$$\left(\hat{\imath}_t - E_t\left[\hat{\pi}_{t+1}\right]\right) - \left(\hat{\imath}_t^* - E_t\left[\hat{\pi}_{t+1}^*\right]\right) = E_t\left[\widehat{rs}_{t+1}\right] - \widehat{rs}_t \tag{2.17}$$

The risk sharing equation (2.11) becomes

$$\widehat{rs}_t = \sigma\left(\hat{c}_t - \hat{c}_t^*\right) + \epsilon_t^* - \epsilon_t \tag{2.18}$$

The good market clearing condition for domestic market (2.9) yields

$$\hat{y}_t = -\theta\tilde{p}_{H,t} + \hat{c}_t + \lambda\left(\theta - \frac{1}{\sigma}\right)\widehat{rs}_t \tag{2.19}$$

The relationship between relative domestic producer price and relative importer price following from (2.2) is

$$0 = (1-\lambda)\,\tilde{p}_{H,t} + \lambda\tilde{p}_{F,t} \tag{2.20}$$

The relationships between relative producer price and inflation and relative importer price and inflation from (2.6) and (2.7) are given respectively by

$$\tilde{p}_{H,t} - \tilde{p}_{H,t-1} = \hat{\pi}_{H,t} - \hat{\pi}_t \tag{2.21}$$

and

$$\tilde{p}_{F,t} - \tilde{p}_{F,t-1} = \hat{\pi}_{F,t} - \hat{\pi}_t \tag{2.22}$$

The LOP gap (2.4) is

$$\hat{\Psi}_t = \widehat{rs}_t - \tilde{p}_{F,t} \tag{2.23}$$

The relationship between real and nominal exchange rate in (2.5) is expressed by the law of motion

$$\Delta\widehat{rs}_t = \Delta\hat{s}_t + \hat{\pi}^*_t - \hat{\pi}_t + \varepsilon_{rs,t} \tag{2.24}$$

where $\varepsilon_{rs,t}$ is an unobservable shock, to capture possible measurement error in the data and to relax the potentially tight cross-equation restrictions in the model.

The log-linearization of the supply side leads to a hybrid NKPC with a non-zero steady state inflation

$$\hat{\pi}_{H,t} = \chi^f E_t\left[\hat{\pi}_{H,t+1}\right] + \chi^b \hat{\pi}_{H,t-1} + \kappa_{mc}\left(\widehat{mc}_t + v_t\right) + \chi^\pi\left(\hat{h}_t - \left(\hat{y}_t - \sigma\hat{c}_t\right)\right) \tag{2.25}$$

where the real marginal cost is

$$\widehat{mc}_t = \eta\hat{y}_t + \sigma\hat{c}_t - (\eta + 1)\, a_t - \tilde{p}_{H,t} \tag{2.26}$$

and

$$\hat{h}_t = \left(1 - \alpha\beta\Pi^{\varepsilon-1}\right)\left(\hat{y}_t - \sigma\hat{c}_t\right) + \left(\alpha\beta\right)\Pi^{\varepsilon-1}E_t\left[\varepsilon\hat{\pi}_{H,t+1} - \hat{\pi}_{t+1} + \hat{h}_{t+1}\right] \quad (2.27)$$

Analogously, the NKPC for imported prices can be log-linearized to obtain

$$\hat{\pi}_{F,t} = \chi_F^f E_t\left[\hat{\pi}_{F,t+1}\right] + \chi_F^b\hat{\pi}_{F,t-1} + \kappa_F\left(\widehat{\Psi}_t + v_t^F\right) + \chi_F^\pi\left(\hat{h}_t^F - \left(\hat{c}_t^F - \sigma\hat{c}_t\right)\right) \quad (2.28)$$

with

$$\hat{h}_t^F = \left(1 - \alpha^F\beta\Pi^{\varepsilon-1}\right)\left(\hat{c}_t^F - \sigma\hat{c}_t\right) + \left(\alpha^F\beta\right)\Pi^{\varepsilon-1}E_t\left[\varepsilon\hat{\pi}_{F,t+1} - \hat{\pi}_{t+1} + \hat{h}_{t+1}^F\right] \quad (2.29)$$

and

$$\hat{c}_{F,t} = \hat{c}_t - \theta\tilde{p}_{F,t} \quad (2.30)$$

where $\hat{c}_{F,t}$ is consumption of imported foreign goods.

The market clearing condition for the large economy is

$$\hat{y}_t^* = \hat{c}_t^* \quad (2.31)$$

The foreign Euler Equation yields

$$\hat{c}_t^* = E_t\left[\hat{c}_{t+1}^*\right] - \frac{1}{\sigma}\left(\hat{\imath}_t^* - E_t\left[\hat{\pi}_{t+1}^*\right] + E_t\left[\Delta\varepsilon_{t+1}^*\right]\right) \tag{2.32}$$

The Phillips curve with a backward looking and non-zero inflation component is identical to the one for a closed economy

$$\hat{\pi}_t^* = \chi_f^* E_t\left[\hat{\pi}_{t+1}^*\right] + \chi_b^*\hat{\pi}_{t-1}^* + \kappa_{mc}^*\left(\widehat{mc}_t^* + v_t^*\right) + \chi_\pi^*\left[\hat{h}_t^* + (\sigma-1)\,\hat{y}_t^*\right] \tag{2.33}$$

where

$$\hat{h}_t^* = \left(1-\alpha\beta\Pi^{\varepsilon-1}\right)\left(\hat{y}_t^* - \sigma\hat{c}_t^*\right) + (\alpha\beta)\,\Pi^{\varepsilon-1}E_t\left[(\varepsilon-1)\,\hat{\pi}_{t+1}^* + \hat{h}_{t+1}\right], \tag{2.34}$$

and the marginal cost is

$$\widehat{mc}_t^* = (\eta+\sigma)\,\hat{y}_t^* - (1+\eta)\,a_t^* \tag{2.35}$$

2.1.3 Monetary policy rules

To close the model, a monetary policy rule needs to be specified. For estimation purposes, it is customary to use a generalized Taylor rule. Analyzing the effect of such a simple rule has some advantages relative to the optimal monetary policy, as it is more likely to be used in practice because it is more easily implemented. Additionally, the parameters are more robust to the model specification than the structural parameters of the optimal rule. Thus, the monetary policy rule takes the form

$$\hat{\imath}_t^* = \rho_i^* \hat{\imath}_{t-1}^* + \phi_\pi^* \hat{\pi}_t^* + \phi_y^* \hat{y}_t^* + \phi_\Delta^* \Delta\hat{\pi}_t^* + \varepsilon_{u,t}^* \tag{2.36}$$

where $\varepsilon_{u,t}^*$ is an exogenous monetary policy shock.

For the small economy, the monetary policy rule is analogous to (2.36). By adjusting the interest rate, the central bank targets CPI inflation, its speed, domestic output growth and the exchange rate. By setting $\phi_S \neq 0$ in the monetary policy rule, I obtain

$$\hat{\imath}_t = \rho_i \hat{\imath}_{t-1} + \phi_\pi \hat{\pi}_t + \phi_y \hat{y}_t + \phi_{\Delta 1} \Delta\hat{\pi}_t + \phi_S \Delta\hat{s}_t + \varepsilon_{u,t} \tag{2.37}$$

Exogenous Disturbances

The model contains seven exogenous shocks that follow autoregressive processes expressed in a log-linearized form. The country-specific TFP for domestic and foreign country are defined respectively by

$$a_t = \rho_a a_{t-1} + \varepsilon_{a,t}$$

$$a_t^* = \rho_{a^*} a_{t-1}^* + \varepsilon_{a,t}^*$$

the preference innovations are given for domestic and foreign consumers respectively by

$$\epsilon_t = \rho_e \epsilon_{t-1} + \varepsilon_{e,t}$$

$$\epsilon_t^* = \rho_{e^*} \epsilon_{t-1}^* + \varepsilon_{e,t}^*$$

Finally, the cost push for domestic producers and for domestic retailers are expressed by

$$v_t = \rho_v v_{t-1} + \varepsilon_{v,t}$$

$$v_t^F = \rho_{v^F} v_{t-1}^F + \varepsilon_{v^F,t}$$

whereas for foreign producers by

$$v_t^* = \rho_{v^*} v_{t-1}^* + \varepsilon_{v,t}^*$$

To summarize, the model exhibits nine structural shocks, of which seven are white noise entering the above AR(1) processes, namely $\varepsilon_{a,t}$, $\varepsilon_{a,t}^*$, $\varepsilon_{e,t}$, $\varepsilon_{e,t}^*$, $\varepsilon_{v,t}$, $\varepsilon_{v^F,t}$, $\varepsilon_{v,t}^*$, and two are exogenous monetary policy shocks, namely $\varepsilon_{u,t}$ and $\varepsilon_{u,t}^*$; plus one measurement error, $\varepsilon_{rs,t}$.

2.2 Estimation and forecasting results

This subsections contains estimation results of the DSGE model presented above and of the identical model assuming a zero steady state inflation. To estimate and to forecast the models I use Dynare version 4.4.[4] The models are estimated in its log-linearized form by using a Bayesian methodology. The large economy is represented by the Euro Area and the small open economy is in turn one of three EEC countries, namely the Czech Republic, Hungary and Poland. The forecasting results are evaluated using predictive distribution and point forecast accuracy.

This section is divided into four parts. First, I give some details about the data I use to estimate the model. In the second part, I discuss the Bayesian methodology and estimation technique. In the third part, I describe my choice of priors in the context of the existing literature on this field. Finally, I illustrate the estimation and forecasting results.

2.2.1 Data and measurement equations

For my empirical analysis and forecasting I gather a dataset of observations on output growth, CPI inflation, PPI inflation, interest rates, exchange rates, and a terms of trade related variable: the relative domestic price. The small open economy is, in turn, one of three EEC countries, namely the Czech Republic, Hungary

[4] The relevant, DYNARE package, which is written for the software MATLAB, is available at http://www.cepremap.cnrs.fr/dynare/.

and Poland. The large economy is represented by the Euro area. The dataset is comparable with those found in the literature, such as Lubik and Schorfheide (2007) or Del Negro and Schorfheide (2009), to name just two. Needless to say, these authors use term of trade composed as a (log-)ratio of export and import price indices. Because of lack of these data for the selected countries, I instead use the relative domestic price that is given by the (log-)ratio between domestic and consumer price index.

The source of the data is the FRED database and the details on each of the particular time series are given in Appendix 2.A. The description of the time series included in the analysis is as follows. The dataset contains quarterly data that are seasonally adjusted using the defaults settings of the X12 filter in Eviews 6. The empirical analysis is based on a sample over the period 1996 to 2013 for the Czech Republic and Poland, and 1998 to 2013 for Hungary. The CPI inflation is constructed as the log difference of the consumer or producer price index multiplied by 100, and the output growth is seasonally adjusted growth rate of GDP in constant prices. Furthermore, the interest rate is an annualized interest rate divided by four so as to be expressed in quarterly terms.[5] The exchange rate is given by the seasonally adjusted real effective exchange rate, computed

[5] The interest rate is given as the 3-month or 90-day rates and yields interbank rates for the Euro Area, the Czech Republic and Poland and as the 3-Month or 90-day rates and yields treasury securities for Hungary. The fact that I cannot use the interbank rate for Hungary is due to missing values in the time series.

as the weighted average of bilateral exchange rates adjusted by relative consumer prices. Finally, the relative domestic price is computed as the logarithm of the ratio between seasonally adjusted domestic and consumer price index.

The data are displayed in Figures 2.1 and 2.2. It is immediate to notice from Figure 2.1 that all countries experienced a strong GDP slow down during the financial crisis. The strong increase in the PPI inflation around 2000 in the Czech Republic can be seen as an adjustment to the inflation base. However, one can see clearly that, in Hungary, inflation was more volatile than in other countries, and that Poland, at the end of the 1990s, experienced large volatility in the output growth. The interest rates of the EECs were much higher in the 1990s and at the beginning of the century than in the Euro Area, whereas nowadays they are stable and low for all the considered countries. Figure 2.2 illustrates the real exchange rate and the terms of trade, here expressed as the relative domestic price. It is already apparent from visual investigation that there is a strong correlation between these two variables. For all the considered countries, the real exchange rate appreciates through time. Additionally, it is worth noting that the choice of these variables is common in the literature. As for the estimation of the NKPC, some authors also add unit labour costs as a proxy for the real marginal costs (for a discussion, see Junicke, 2017). However, most of the empirical papers take the marginal costs as a latent variable. For this reason, as Schorfheide (2008)

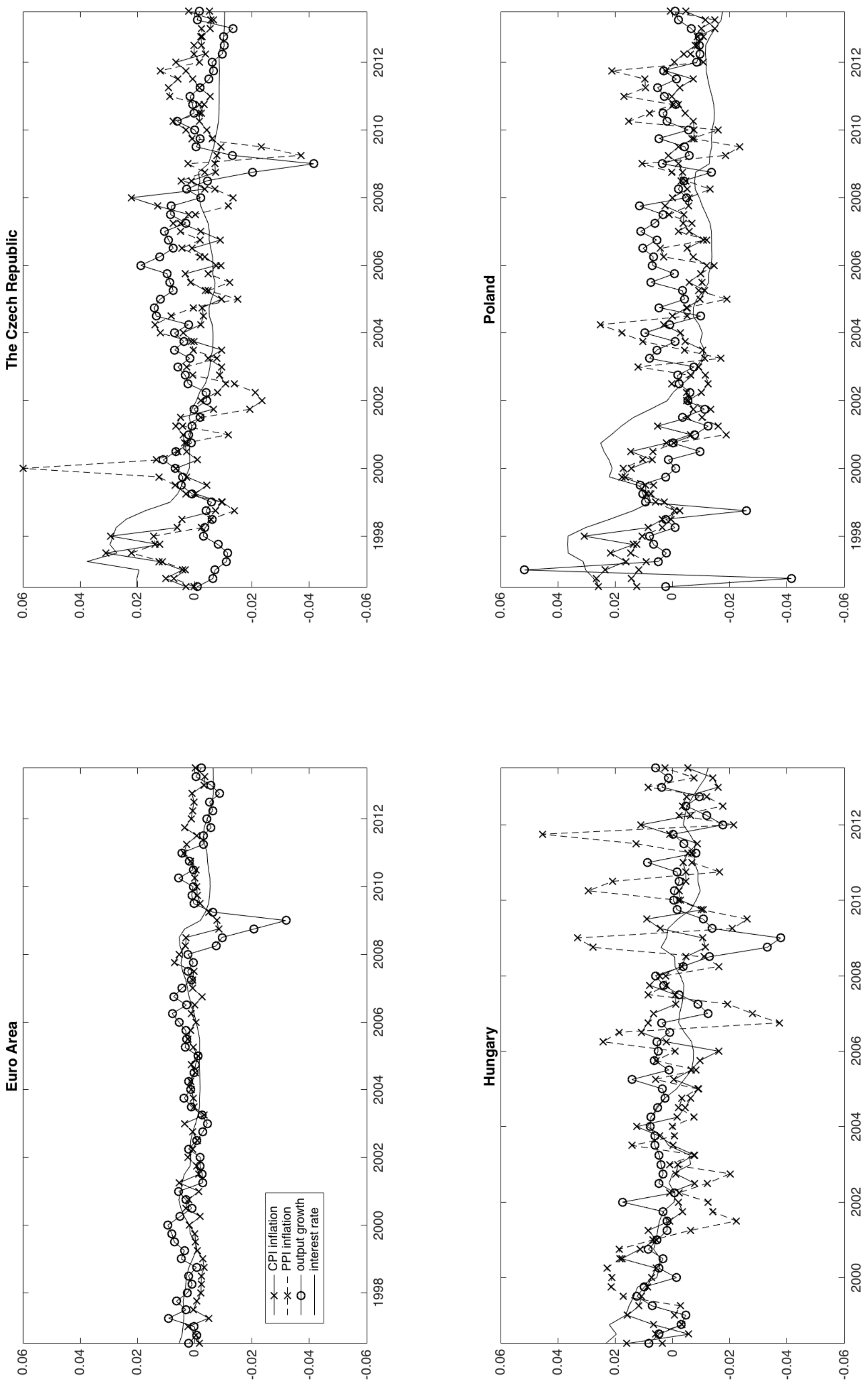

Figure 2.1. Inflation, GDP growth and interest rate time patterns in the Euro Area and in the three EECs.

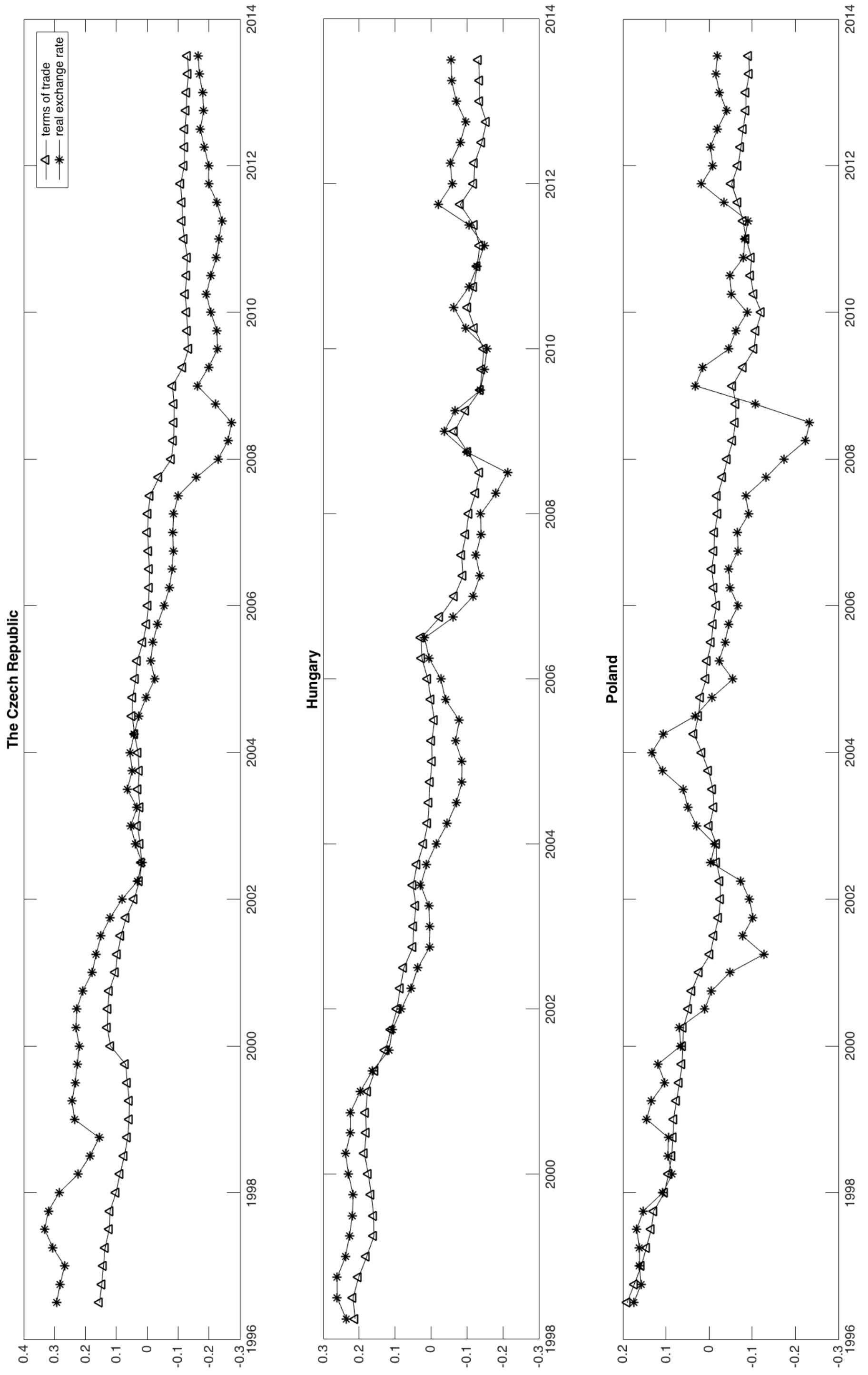

Figure 2.2. Terms of trade and exchange rate time patterns in the three EECs.

describes, the estimation results on the NKPC parameters may vary significantly.

To pinpoint the relationship between the observed time series and the model variables, it is useful to rewrite the model in the state space form. The likelihood function is then calculated using a Kalman filter; see, *e.g.*, Hamilton (1994) for details. The estimated model consists of a set of equilibrium equations that are log-linearized, and the variables are expressed in terms of the deviation from their respective steady state levels, both for the small and the large economy. The solution of the system of equations can be expressed in the form

$$s_t = \Phi_1(\theta)\, s_{t-1} + \Phi_e(\theta)\, e_t \tag{2.38}$$

known as the transition equation. Here s_t is a vector of state variables, given by

$$s_t = [\hat{\pi}_t, \hat{y}_{t,t}\,, \widehat{rs}_{t,}\tilde{p}_{H,t}, \pi^*_t, \hat{y}^*_t, \hat{\imath}^*_t]$$

and $e_t \sim N(0, Q(\theta))$ is an i.i.d. vector of the exogenous shocks described in the previous section. The matrix $\Phi(\theta)$ collects the structural parameters of the model. The measurement equation, also called observation equation, relates the model variables s_t to the observable variables in vector y_t in such a way that

$$y_t = A(\theta) + Bs_t + v_t \tag{2.39}$$

where $A(\theta)$ is a vector containing the trend of the observed variables, B is the coefficients matrix, which relates the observables with the model states, and v_t is vector of the measurement errors. To prevent the problem of stochastic singularity, it must hold that the number of time series is lower than number of shocks. The frequency of the observed variables is one quarter, it and is also assumed that one time period in the model corresponds to one quarter of a year.

Most of the observed time series are manipulated in such a way that they correspond directly to the relevant variables in the model. Using a loglinearized model in Dynare, the steady state is always set to be zero. Therefore, there are two options. First, one can detrend the observed time series in such a way that they oscillate around a zero mean. Second, one can add the observed steady state value to the (theoretical) variable. Here, the value of steady state inflation is taken from the data long run trend.

In more details, the observation equations can be written as follows. For inflation, it holds[6]

$$\pi_t^{OBS} = [\hat{\pi}_t + \log(\Pi)]$$

[6]The data are manipulated in such a way that it holds $\pi_t^{OBS} = \log\left(P_t^{data}/P_{t-1}^{data}\right)$. For the variable inflation in model, it holds $\hat{\pi}_t = \log(\Pi_t) - \log(\Pi)$, with $\Pi_t = P_t/P_{t-1}$. So that the observation equation for inflation can be written as $\pi_t^{OBS} = \hat{\pi}_t + \log \Pi$.

For output, whose time series is given as a first difference, it holds

$$y_t^{OBS} = (\hat{y}_t - \hat{y}_{t-1})$$

For the interest rate, which is not detrended either, once must add the level of steady state $\bar{R} = \Pi/\beta$ to obtain

$$i_t^{OBS} = \hat{\imath}_t + \log\left(\bar{R}\right)$$

The relationship between observed and modelled relative domestic price can be simply expressed by

$$p_{H,t}^{OBS} = \tilde{p}_{H,t}$$

and for the real exchange rate, detrended and in log terms, it holds

$$rs_t^{OBS} = \widehat{rs}$$

First three equations can be also used for the foreign economy, associated with the Euro Area data.

2.2.2 Choice of priors

Selecting appropriate priors before the Bayesian inference is an important task. The result of the estimation may in fact vary significantly depending on how loose the priors are set. If just a small sample of data is available, as it is the case in this chapter, a prior distribution is additional information that enables more stability in the optimization algorithm. Thus, my strategy is to set the priors tight when a sound explanation for the values of priors exists and/or such values are economically interpretable. The priors in this chapter are set in line with the existing literature. For the Euro Area, the selection of the prior distribution follows closely Smets and Wouters (2003, 2007), and is represented in Table 2.1; for the SOEs, the priors are consistent with those in Lubik and Schorfheide (2007) and Marcellino and Rychalovska (2014), and are presented in Table 2.2.

For some parameters, dogmatic priors are imposed. There are two reasons. First, in a model without capital and investment, it is difficult to estimate parameters that are connected to the capital market, such as β. The second reason is known in the literature as the identification problem (see, *e.g.*, Rabanal and Rubio-Ramírez, 2015): it is impossible to estimate certain parameters at the same time, for instance the probability of changing the price α and the price mark up ε, because it is not possible to identify them simultaneously.

Parameter	Distribution	Mean	Standard error
$\sigma\left(\varepsilon_a^*\right)$	Inverse Gamma	1	2
$\sigma\left(\varepsilon_e^*\right)$	Inverse Gamma	1	2
$\sigma\left(\varepsilon_v^*\right)$	Inverse Gamma	1	2
$\sigma\left(\varepsilon_u^*\right)$	Inverse Gamma	1	2
ρ_a^*	Beta	0.8	0.1
ρ_e^*	Beta	0.8	0.1
ρ_v^*	Beta	0.8	0.1
σ^*	Gamma	2	0.4
η^*	Gamma	3	0.9
α^*	Beta	0.7	0.2
ω^*	Beta	0.5	0.1
ϕ_π^*	Gamma	1.5	0.2
ϕ_y^*	Gamma	0.125	0.05
$\phi_{\Delta 1}^*$	Gamma	0.3	0.1
$\phi_{\Delta 2}^*$	Gamma	0.0625	0.05
$\phi_{\Delta y}^*$	Gamma	0.0625	0.05
ρ_i^*	Beta	0.8	0.1

Table 2.1. Prior distribution for the large economy

The average inflation of the estimated sample for the Euro Area is set to $\Pi = 1.005$, corresponding to a 2% trend inflation a year.[7] For the rest of the countries, the trend inflation prior is slightly higher, $\Pi = 1.0076$, $\Pi = 1.0142$ and $\Pi = 1.0121$, for the Czech Republic Hungary and Poland, respectively. The priors for the SOE importer NKPC parameter are set analogously to the producer NKPC. The degree of openness λ is set to 0.6 for the Czech Republic, corresponding to the average Import/GDP ratio over the data sample. For Hungary and Poland, it is set to be 0.7 and 0.36, respectively.

The priors for the parameters of the utility function σ and η are taken from the relevant literature. For parameters that are restricted to the interval $(0, 1)$, I use a Beta distribution. Non-negative parameters are then Gamma distributed. As for the autoregressive parameters of the shocks, I use a Beta distribution with a mean of 0.8 and a standard deviation of 0.1. The variances of the shocks are inverse gamma, with unity prior mean and with two degrees of freedom. The standard errors are set such that the domain covers a reasonable range of parameter values, hence their prior is rather loose.

The priors for the interest rate rule coefficients have rather wide confidence intervals. They are distributed around a mean given by the Taylor rule, following Lubik and Schorfheide (2005). Additionally, the prior distribution for the parame-

[7] This average CPI inflation also corresponds with the inflation target of the ECB.

Parameter	Distribution	Mean	Standard error
$\sigma\left(\varepsilon_a\right)$	Inverse Gamma	1	2
$\sigma\left(\varepsilon_e\right)$	Inverse Gamma	1	2
$\sigma\left(\varepsilon_u\right)$	Inverse Gamma	1	2
$\sigma\left(\varepsilon_v\right)$	Inverse Gamma	1	2
$\sigma\left(\varepsilon_{v^F}\right)$	Inverse Gamma	1	2
$\sigma\left(\varepsilon_{rs}\right)$	Inverse Gamma	1	2
ρ_a	Beta	0.8	0.1
ρ_e	Beta	0.8	0.1
ρ_v	Beta	0.8	0.1
ρ_{v^F}	Beta	0.8	0.1
σ	Gamma	2	0.5
η	Gamma	2	1.5
α	Beta	0.7	0.2
α^F	Beta	0.7	0.2
ω	Beta	0.5	0.1
ω^F	Beta	0.5	0.1
ρ_i	Beta	0.8	0.1
ϕ_π	Gamma	1.5	0.1
ϕ_y	Gamma	0.125	0.05
$\phi_{\Delta 1}$	Gamma	0.3	0.1
$\phi_{\Delta 2}$	Gamma	0.0625	0.05
$\phi_{\Delta y}$	Gamma	0.0625	0.05
ϕ_S	Gamma	0.3	0.1

Table 2.2. Prior distribution for the small open economy

ter ϕ_π has a lower bound of one, to satisfy the Taylor principle. Priors for the rest of the parameters in the monetary policy rule are Gamma distributed, with mean and standard error as those chosen by Smets and Wouters (2003) and Lubik and Schorfheide (2005). It is worth mentioning that most of the parameters are not imposed to be the same for all countries, but it is merely assumed that they have identical priors.

2.2.3 Methodology

This section illustrates the estimation of the model, and is divided into three parts. First, I discuss the Bayesian methodology and estimation technique. Then, I turn to describe the forecast methodology and some tools that I use to evaluate the forecasting accuracy. The next subsection presents the forecasting results.

To proceed with the estimation and the subsequent forecasting exercise, I choose Bayesian inference over maximum likelihood estimation. This allows to incorporate a prior distribution that, as discussed earlier, enables introducing additional general information about subjective beliefs on the parameter distribution, or information coming from previous econometric and theoretical studies.

Suppose that the aim is to draw a sample from a target density $\pi(\Phi)$. Note that Φ is a $(K \times 1)$ vector of parameters of interest. The target density is a posterior distribution, which is too complex to allow for a direct sample. Therefore an indirect method is needed. The steps describing a random walk Metropolis-Hastings algorithm are the following.

1. Set a prior distribution for each parameter $p(\Phi)$.

2. Find the mode of the posterior distribution $\pi(\Phi)$ via numerical maximization. Denote the estimates of the parameters at the mode by $\Phi^{\max}$, and their covariance matrix, which is the inverse Hessian matrix, by $H^{\max}$.

3. To approximate $\pi(\Phi)$, the following algorithm is used.

   (a) Specify a candidate density $q\left(\Phi^{G+1}/\Phi^{G}\right)$, where G is an index of draws.

   (b) Set the initial estimates of the parameters Φ^{G} with $G = 0$.

   (c) Generate a candidate value Φ^{G+1} from the candidate density. I use a random walk version of this algorithm with the candidate density specified as a random walk,

   $$\Phi^{G+1} = \Phi^{G} + e$$

   where e is a K-vector random walk with a normal distribution

   $$e \sim N(0, \Sigma)$$

   (d) Compute the acceptance probability. The candidate Φ^{G+1} is accepted with probability α, given by

$$\alpha = \min \left(\frac{\pi \left(\Phi^{G+1} \right) / q \left(\Phi^{G+1} / \Phi^{G} \right)}{\pi \left(\Phi^{G} \right) / q \left(\Phi^{G} / \Phi^{G-1} \right)} ; 1 \right)$$

where the numerator is the target density evaluated at the new draw of the parameters $\pi \left(\Phi^{G+1} \right)$ relative to the candidate density evaluated at the new draw parameters $q \left(\Phi^{G+1} / \Phi^{G} \right)$, and the denominator is the same expression evaluated at the previous draw of the parameters. Using a random walk version together with the fact that the normal distribution is symmetric, the acceptance probability simplifies to

$$\alpha = \min \left(\frac{\pi \left(\Phi^{G+1} \right)}{\pi \left(\Phi^{G} \right)} ; 1 \right)$$

Step 3 is repeated M times. The first $(M - J)$ iterations are discarded. The last J draws are instead retained to estimate the posterior marginal distribution. For the results, I use four chains of $M = 500,000$ draws, each starting from a different value. From each chain, the last $J = 0.55 \times M$ draws are used to approximate the empirical distribution of the parameters.

Using the Metropolis-Hastings algorithm, the acceptance rate depends on the variance Σ, which is set manually. It holds that the higher the variance, the more volatile the drawings. Therefore, a lower acceptance is to be expected in this case. Conversely, if Σ is set too low, the volatility of the drawings is low as well. Therefore, the estimation of the parameters is likely to be close to the prior. Drawing a random number u from a uniform distribution $u \sim U(0,1)$, it holds that the candidate Φ^{G+1} is accepted if $\alpha > u$, otherwise it is rejected.

The acceptance rate, given by the ratio between the accepted draws and the total number of draws, should lie between 20% and 40%. Some researchers are more specific and suggest that, for multivariate estimations, the acceptance rate should optimally be set to approximately 23%. The convergence of the chains is checked according to the Brooks and Gelman (1998) convergence diagnostic.

To produce a forecast, we generate J draws from the posterior predictive distribution of $Y_{T+1:T+H}$, given $Y_{1:T}$. The algorithm is as follows, for $j = 1...J$.[8]

1. Draw $\left(\theta^{j}, s_T^{(j)}\right)$ from the posterior distribution $\pi(\theta, s_T|Y_{1:T})$.
2. Draw from $\pi\left(s_{T+1:T+H}|\theta^{(j)}, s_T^{(j)}\right)$ as follows.
   (a) Draw the structural shock innovation $e_{i,T+1:T+H}^{(j)} \sim N\left(0, \sigma_i^{2(j)}\right)$ for a structural shock i.

[8] For more informarmation, see Diebold *et al.* (2016).

(b) Starting from $s_T^{(j)}$, iterate the transition equation (2.38) forward

$$\begin{aligned} s_t^{(j)} &= \Phi_1\left(\theta^{(j)}\right) s_{t-1}^{(j)} + \Phi_e\left(\theta^{(j)}\right) e_t^{(j)} \\ \text{for } t &= T+1, ..., T+H \end{aligned}$$

3. Compute the sequence $Y_{T+1:T+H}^{(j)}$, using measurement equation (2.39)

$$\begin{aligned} y_t^{(j)} &= A\left(\theta^{(j)}\right) + B s_t^{(j)} + v_t^{(j)} \\ \text{for } t &= T+1, ..., T+H \end{aligned}$$

This paper compares the forecasting performance of rolling and recursive forecasting scheme of the model assuming zero steady state inflation and non zero steady state inflation. Using a recursive scheme, for each date t the data set is extended by one period to $t+1$ observations, and the parameters are reestimated accordingly. Thus, the number of observations increase until the last in-sample estimation. On the contrary, using a rolling scheme, the number of observations does not change: the scheme uses a rolling windows of $R+i-j$ observations to re-estimate the parameters. In the presence of structural breaks, an advantage may arise in using a rolling scheme.

In what follows, I present the results on point forecasts for both a recursive and a rolling schemes. I compare the root mean square error (RMSE), which measures the deviation of the predicted values from the observed ones, for the small open economy output (growth), inflation, interest rate and exchange rate

$$RMSE\left(i|h\right) = \sqrt{\frac{1}{P-h}\sum_{t=R}^{R+P-h}\left(y_{i,t+h} - \hat{y}_{i,t+h|t}\right)^2}$$

where R denotes the starting point of the forecast evaluation sample, and P is the number of forecast origins.

For the Czech Republic, 69 observations are available. I use $R = 40$ as the starting point of the in-sample estimation, *i.e.*, the first estimation is run over the period from April 1996 to October 2003. As a result, the first forecast by both the rolling and the recursive scheme is done for period from January 2004 to October 2005. (Notice that data are collected in January, April, July and October.) For the recursive forecast, the sample is expanding, and the last one contains all 61 observations, from April 1996 to July 2011. For the rolling forecast, the last sample that produces forecast directly comparable with real data incorporates observations from April 2004 to July 2011.

2.2.4 Results

I begin the section by giving the parameter estimates for each country. Then I continue my analysis by presenting the forecasts, and discussing their accuracy using the RMSE computation.

Parameter estimates

The Bayesian estimated posterior distribution for the Euro Area is reported in Table 2.3. The table displays the mode and standard error resulting from the posterior maximization. It also details the estimation results obtained through the Metropolis-Hastings algorithm, such as the posterior mean and the 90% posterior probability interval for both the estimated parameters and the standard deviation of shocks.

For all values, the highest posterior density intervals suggest that the estimated parameters are significantly different from zero. The parameter estimates are in line with the existing literature. Note that the two models deliver similar parameter estimates. Table 2.3 also shows that price stickiness in the Euro Area is seemingly very high, with the parameter α around 0.9, while the backward looking parameter ω takes a value around 0.3, which is lower than that assumed in the prior distribution. The values are robust and lie in the confidence interval for all estimations. With the exception of the technology shock, whose value is a bit lower

Parameter	Mode	S.D.	5%	Mean	95%
$\sigma\left(\varepsilon_a^*\right)$	0.2477	0.0074	0.2791	0.2985	0.3199
$\sigma\left(\varepsilon_e^*\right)$	0.1733	0.0049	0.0049	0.1651	0.1651
$\sigma\left(\varepsilon_u^*\right)$	0.1290	0.0052	0.0052	0.1323	0.1323
$\sigma\left(\varepsilon_v^*\right)$	0.2817	0.0271	0.2157	0.2670	0.3105
ρ_a^*	0.6443	0.0140	0.6051	0.6214	0.6385
ρ_e^*	0.9808	0.0250	0.9756	0.9793	0.9830
ρ_v^*	0.8876	0.0078	0.8827	0.8972	0.8972
σ^*	1.9731	0.0483	1.9614	2.0195	2.0933
η^*	4.8849	0.0883	4.8186	4.9439	5.0928
α^*	0.9561	0.0276	0.9551	0.9578	0.9603
ω^*	0.3437	0.0083	0.3375	0.3456	0.3535
ϕ_π^*	1.0760	0.0219	1.0522	1.0836	1.1115
ϕ_y^*	0.1285	0.0097	0.1223	0.1398	0.1602
$\phi_{\Delta 1}^*$	0.1526	0.0119	0.1416	0.1565	0.1732
$\phi_{\Delta 2}^*$	0.0221	0.0042	0.0134	0.0225	0.0312
$\phi_{\Delta y}^*$	0.0025	0.0017	0.0018	0.0055	0.0105
ρ_i^*	0.4536	0.0110	0.4287	0.4444	0.4610

Table 2.3. Parameter estimation results for the Euro Area

Parameter	Mode	S.D.	5%	Mean	95%
$\sigma(\varepsilon_a)$	0.1290	0.0114	0.1290	0.1376	0.1471
$\sigma(\varepsilon_e)$	0.1694	0.0063	0.1476	0.1754	0.1946
$\sigma(\varepsilon_u)$	0.1290	0.0045	0.1290	0.1343	0.1412
$\sigma(\varepsilon_v)$	0.2733	0.0461	0.2506	0.2953	0.3348
$\sigma(\varepsilon_{v^F})$	0.2170	0.0737	0.1629	0.2407	0.3100
ρ_a	0.6832	0.0065	0.6667	0.6874	0.7072
ρ_e	0.9699	0.0118	0.9489	0.9639	0.9783
ρ_v	0.8878	0.0097	0.8593	0.8820	0.9044
ρ_{v^F}	0.8790	0.0110	0.8769	0.8971	0.9190
σ	2.2826	0.0289	2.2614	2.3423	2.4288
η	4.2722	0.0534	4.0785	4.2398	4.3805
α	0.5164	0.0089	0.4960	0.5215	0.5380
α^F	0.7803	0.0089	0.7442	0.7933	0.8361
ω	0.5116	0.0089	0.4799	0.5188	0.5576
ω^F	0.4276	0.0040	0.4187	0.4333	0.4494
ϕ_π	1.4709	0.0115	1.4492	1.4764	1.5075
ϕ_y	1.4709	0.0057	0.0573	0.0709	0.0870
ϕ_S	0.4921	0.0110	0.4793	0.4958	0.5185
$\phi_{\Delta 1}$	0.5210	0.0152	0.5109	0.5184	0.5260
$\phi_{\Delta 2}$	0.0516	0.0037	0.0461	0.0528	0.0585
$\phi_{\Delta y}$	0.0677	0.0031	0.0512	0.0666	0.0782
ρ_i	0.7511	0.0078	0.7334	0.7498	0.7704

Table 2.4. Parameter estimation results for the Czech Republic

Parameter	Mode	S.D.	5%	Mean	95%
$\sigma\left(\varepsilon_a\right)$	0.1290	0.0088	0.1290	0.1417	0.1568
$\sigma\left(\varepsilon_e\right)$	0.1373	0.0074	0.1295	0.1442	0.1561
$\sigma\left(\varepsilon_u\right)$	0.1290	0.0060	0.1290	0.1354	0.1427
$\sigma\left(\varepsilon_v\right)$	0.3140	0.0423	0.2661	0.3484	0.4367
$\sigma\left(\varepsilon_{v^F}\right)$	0.2866	0.0694	0.2058	0.3446	0.5080
ρ_a	0.7388	0.0148	0.6943	0.7217	0.7494
ρ_e	0.9523	0.0088	0.9382	0.9590	0.9808
ρ_v	0.9232	0.0121	0.8980	0.9194	0.9415
ρ_{v^F}	0.9123	0.0168	0.8669	0.9056	0.9350
σ	2.9214	0.0311	3.0405	3.1438	3.2439
η	4.9641	0.1240	4.8868	5.2103	5.6061
α	0.5818	0.0262	0.6048	0.6452	0.6811
α^F	0.7864	0.0262	0.7246	0.7890	0.8460
ω	0.4122	0.0317	0.3300	0.3975	0.4837
ω^F	0.5056	0.0210	0.4991	0.5220	0.5463
ϕ_π	1.1651	0.0374	1.0468	1.1096	1.1697
ϕ_y	0.0803	0.0061	0.0517	0.0689	0.0876
ϕ_S	0.5493	0.0179	0.5100	0.5400	0.5726
$\phi_{\Delta 1}$	0.4611	0.0144	0.4523	0.4931	0.5286
$\phi_{\Delta 2}$	0.0023	0.0105	0.0000	0.0079	0.0147
$\phi_{\Delta y}$	0.0787	0.0046	0.0600	0.0884	0.1174
ρ_i	0.7454	0.0176	0.7153	0.7558	0.8075

Table 2.5. Parameter estimation results for Hungary

Parameter	Mode	S.D.	10%	Mean	90%
$\sigma(\varepsilon_a)$	0.2115	0.0077	0.1865	0.2287	0.2287
$\sigma(\varepsilon_e)$	0.1775	0.0121	0.1504	0.1801	0.2094
$\sigma(\varepsilon_u)$	0.1290	0.0055	0.1290	0.1344	0.1421
$\sigma(\varepsilon_v)$	0.3155	0.0398	0.2471	0.3423	0.4507
$\sigma(\varepsilon_{v^F})$	0.3424	0.0706	0.3229	0.4824	0.6446
ρ_a	0.3441	0.0311	0.2367	0.2619	0.2973
ρ_e	0.8805	0.0112	0.8527	0.8748	0.8982
ρ_v	0.7102	0.0333	0.6207	0.6600	0.6933
ρ_{v^F}	0.8546	0.0137	0.7337	0.7748	0.8102
σ	2.5519	0.0526	2.4134	2.5154	2.5932
η	7.0782	0.4273	6.6692	7.2806	8.0326
α	0.8416	0.0654	0.8233	0.8395	0.8554
α^F	0.7298	0.0277	0.6851	0.7492	0.8554
ω	0.6238	0.0081	0.6258	0.6465	0.6757
ω^F	0.4963	0.0162	0.4715	0.5086	0.5477
ϕ_π	1.8396	0.0174	1.9067	1.9865	2.1210
ϕ_y	0.1367	0.0046	0.1667	0.1872	0.2043
ϕ_S	0.3858	0.0089	0.3807	0.4191	0.4666
$\phi_{\Delta 1}$	0.4022	0.0073	0.2727	0.3206	0.3780
$\phi_{\Delta 2}$	0.2259	0.0159	0.1834	0.2092	0.2422
$\phi_{\Delta y}$	0.3124	0.0109	0.2824	0.2982	0.3187
ρ_i	0.6514	0.0171	0.6107	0.6779	0.7462

Table 2.6. Parameter estimation results for Poland

than 0.83, the value estimated by Smets and Wouters (2003), all autoregressive parameters for the shocks are estimated to be higher than 0.8, the value assumed in the prior distribution. Moreover, the monetary policy rules parameters are very robust, and lie in the confidence interval given in Table 2.3. These parameters are all consistent with the values found in the literature.

The estimated parameters for each EEC are illustrated in Table 2.4 to Table 2.6. They are similar across three countries. The mode for the backward looking component for producer inflation lies between 0.4 and 0.6 for all countries. The price stickiness parameter α is lower in the Czech Republic and Hungary, around 0.5, suggesting that small open economies like this have to be more flexible. The parameter is higher for Poland (0.8) comparable with the Euro Area estimates. For all countries, the results are robust and fully in line with the existing literature.

Forecasting Results

In this section, I analyze the forecasting performance of the models discussed in the previous section by computing the root square mean error (RMSE) score for both the model with non zero steady state inflation and the one with zero steady state inflation. The RMSE scores of the two models are illustrated in Figures 2.3-2.5, one for each EEC. In every figure, the top panel displays the RMSE resulting from the rolling scheme forecast, whereas the bottom panel the RMSE resulting from the recursive scheme forecast.

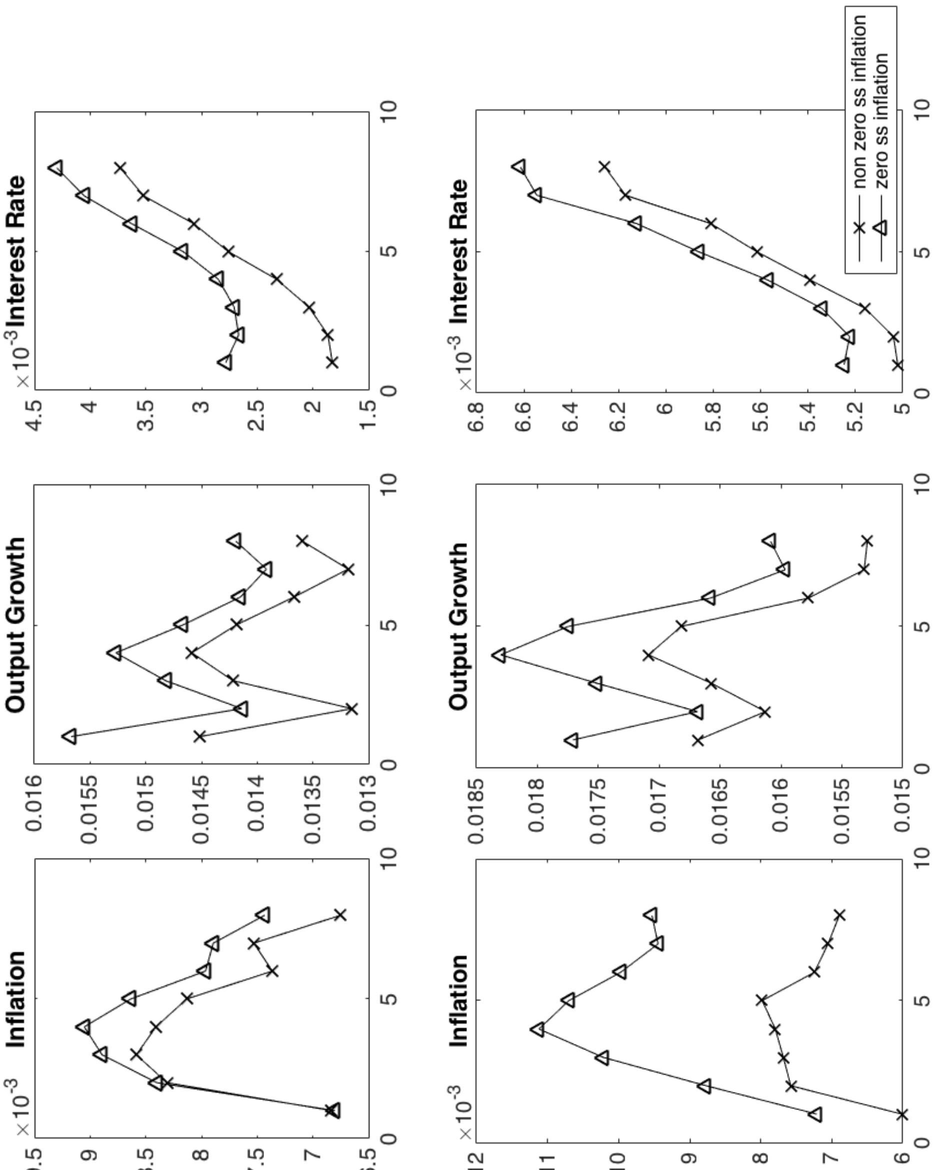

Figure 2.3. Estimation RMSE for the Czech Republic

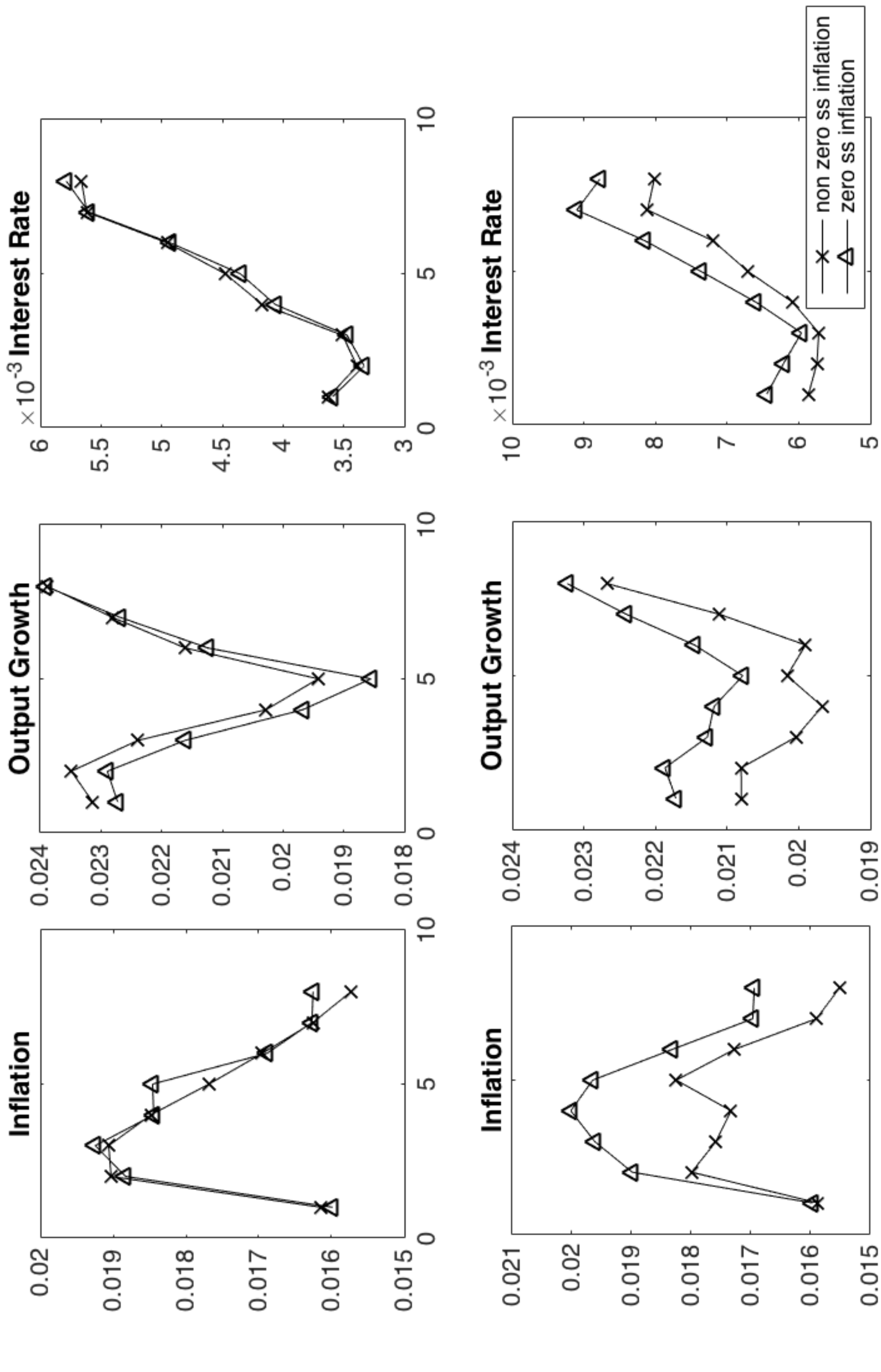

Figure 2.4. Estimation RMSE for Hungary

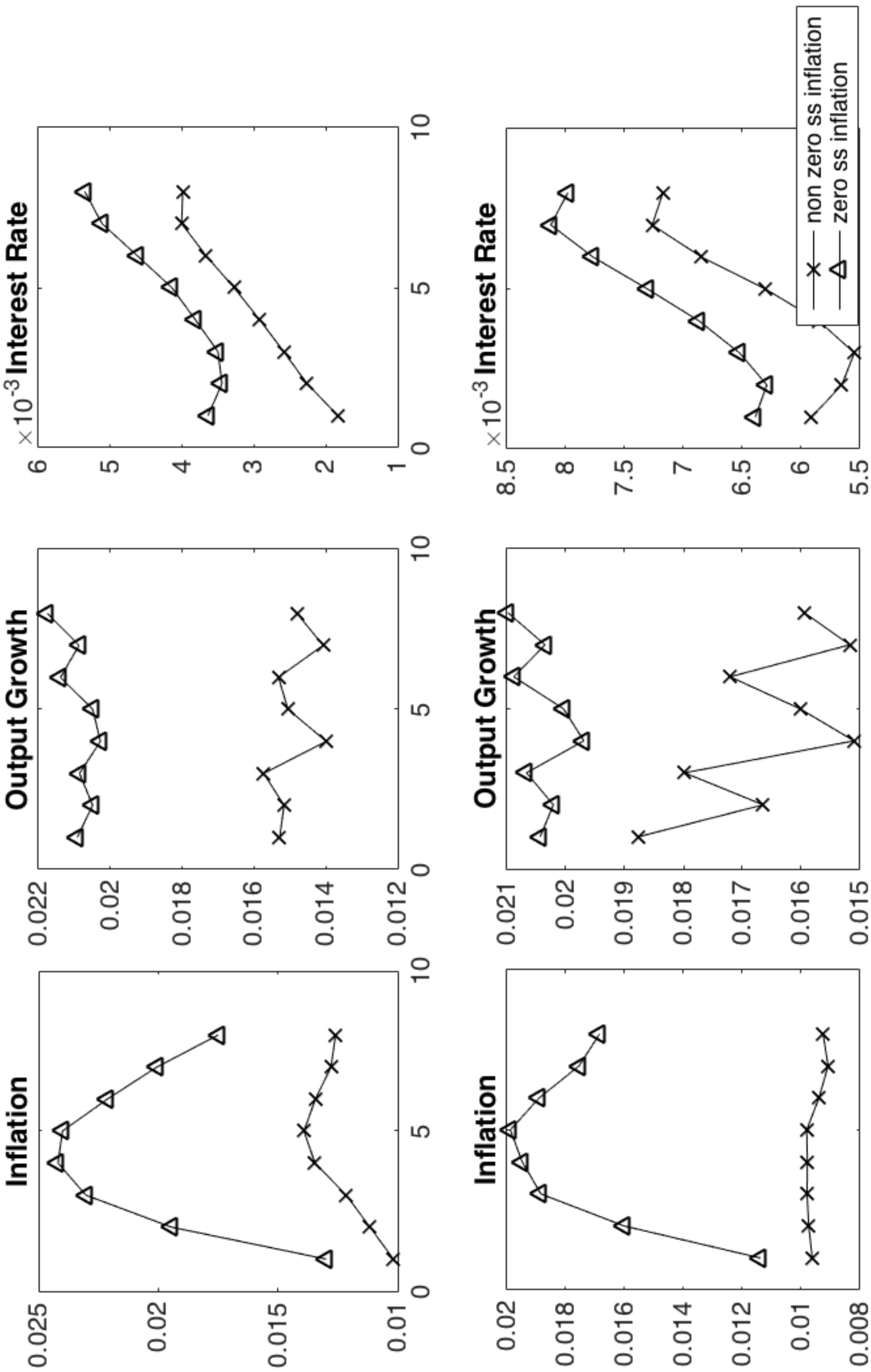

Figure 2.5. Estimation RMSE for Poland

Each graph compares the RMSEs delivered by the two models over a horizon of eight periods. Generally, it emerges that the model with zero steady state inflation never performs better than the one with non zero steady state inflation, in the sense that the forecast given by the first model is never more precise than that delivered by the second model. This holds for each of the three EECs, and using both the rolling and the recursive scheme. Comparing the RMSE forecast of the DSGE model with non zero steady state with its benchmark, it is clear that, in both the cases of rolling and recursive forecasting, the fist one delivers more precise forecast in all cases.

The forecast improvement that the assumption of non-zero steady state inflation delivers is however, more visible in recursive forecast. Thus, *e.g.*, forecasting using rolling scheme inflation delivers similar error in both models at the shorter horizons. However, the spread increases at longer horizons. For output growth, the difference between the model is instead visible even at the shortest horizon, and the RMSE difference remains similar at all horizon lengths. This also holds in the case of recursive forecast for output, but in the case of inflation, the gap between these two models is much more significant by using the recursive scheme.

2.3 Concluding remarks

This chapter has analyzed the forecasting performance of a two-country DSGE model with non zero steady state inflation. First, I have developed a small-scale DSGE model similar to Lubik and Schorfheide (2007), with a micro-founded Phillips curve. I have assumed imperfect pass-through and non-unit intratemporal elasticity of substitution between domestic and foreign goods. I have log-linearized the model around a steady state with non-zero inflation.

I have carried out Bayesian inference, using a Metropolis-Hastings sampling approach, to measure the performance of this model against the Euro area data and. In order to study the model for the SOE, I have used the data of three EECs, namely the Czech Republic, Hungary and Poland. I have shown that the forecasting performance of this model is higher than assuming a benchmark model with zero steady state inflation for all the three countries. This result has been even more remarkable when a recursive forecasting scheme has been used.

2.A Data

The empirical estimation is based on a quarterly data sample over the periods 1996:1 to 2013:3 for Euro Area and Poland, 1996:2 to 2013:3 for the Czech Republic, and 1998:1 to 2013:3 for Hungary.

The FRED database was used as a source for following time series.

- For the Euro Area (17 countries)

  - Consumer Price Index: Harmonized Prices: Total All Items (CPHPTT01EZM661N)
  - Gross Domestic Product by Expenditure, Constant Prices: Total GDP (NAEXKP01EZQ657S)
  - 3-Month or 90-day Rates and Yields: Interbank Rates (IR3TIB01EZQ156N)

- For the Czech Republic

  - Consumer Price Index: All Items (Harmonized CPI) (CZECPIHICQINMEI)
  - Producer Prices Index: Economic Activities: Total Manufacturing (PIEAMP01CZQ661N)
  - Gross Domestic Product by Expenditure, Constant Prices: Total GDP (NAEXKP01CZQ657S)
  - 3-Month or 90-day Rates and Yields: Interbank Rates (IR3TIB01CZQ156N)

  - Real Broad Effective Exchange Rate

    (RBCZBIS)

- For Hungary

  - Consumer Price Index: All Items (Harmonized CPI)

    (HUNCPIHICQINMEI)

  - Producer Prices Index: Economic Activities: Total Manufacturing

    (PIEAMP01HUQ661N)

  - Gross Domestic Product by Expenditure, Constant Prices: Total GDP

    (NAEXKP01HUQ657S)

  - 3-Month or 90-day Rates and Yields: Treasury Securities

    (IR3TTS01HUM156N)

  - Real Broad Effective Exchange Rate

    (RBHUBIS)

- For Poland

  - Consumer Price Index: All Items (Harmonized CPI)

    (POLCPIHICQINMEI)

- Producer Prices Index: Economic Activities: Total Manufacturing (PIEAMP01PLQ661N)
- Gross Domestic Product by Expenditure, Constant Prices: Total GDP (NAEXKP01PLQ657S)
- 3-Month or 90-day Rates and Yields: Interbank Rates (IR3TIB01PLQ156N)
- Real Broad Effective Exchange Rate (RBPLBIS)

Chapter 3

Forecasting with autoregressive models

Does identifying changes in inflation and output dynamics in Eastern European Countries (EECs) help in improving the forecasts of these variables? Of course, if the structure of the economy has changed, a forecasting model that can account for such changes it should be better suited and should deliver better forecasts. That is the reason why models with time varying components gained recently so much attention in the economic literature. Models that incorporate a change in parameters, unlike those with fixed coefficients, are capable to detect structural changes in the variables, and thus they should be able to increase the accuracy of macroeconomic forecasting. However, one must bear in mind that a richer model structure, implying a higher number of parameters, increases the estimation errors and reduces the forecast accuracy.

In this study, we aim to investigate the forecasting performance of models that allow for changes in their parameters relatively to the fixed coefficient model. In order to perform my analysis, I choose three central European countries, namely the Czech Republic, Hungary and Poland. As I discussed in the first chapter of the thesis, I believe that choosing the EECs for our analysis is an interesting case study because of their historical development, an it represents a challenge for the limited amount of available data.

Limited data availability, as it is the case for the EEC, is in fact a well known issue in the forecasting literature. One has to be particularly careful when only short time series are involved, since the model has to be restricted in both the number of variables and the time structure of the model. To assess whether modeling structural change increases the accuracy of macroeconomic forecasts, we aim to investigate the performance of models with of time-varying parameters and covariance matrix, and this only exacerbate the issues just outlined. Nevertheless, filling the gap in the applied forecasting literature regarding countries with only relatively short time series available is an important task, and beginning with the state-of-the-art forecasting models seems the natural way to proceed.

This paper is related to the large literature, reviewed in the first chapter of the thesis, that analyses the forecasting performance using Bayesian technique. My work mainly relates to three sub-groups of this literature. First, from the methodology perspective I follow the work from Primiceri (2005). I develop a time varying structural vector autoregression (SVAR), where time variation are taken into account in both the coefficients and the covariance matrix of the innovations. This is motivated by the objective of distinguishing between changes in the typical size of the exogenous innovations and changes in the transmission mechanism.

Second, regarding the model application, our paper is closely related to D'Agostino *et al.* (2013) and Barnett *et al.* (2014). These authors study the accuracy of the forecasting model with structural change. Using U.S. data, D'Agostino *et al.* (2013) show that a forecasting model that accounts for structural change is able to deliver a better forecast in comparison to several alternative models, including a VAR with fixed coefficients. Barnett *et al.* (2014) compare a large variety of different models with and without time-varying parameters. Using U.K. data, they find that models with time-varying parameters lead to a significant improvement in forecasting performance.

In terms of environment with a severely limited number of observation, the chapter also relates to Mandalinci (2015), who compares different models using emerging market data with short time series. His results indicate that the forecasting performance of different models change notably both across time and countries, though if one accounts for stochastic volatility, then the models perform significantly better.

Third, the contribution in this chapter relates to a recent strand of the literature featuring empirical applications of time-varying models on Eastern European countries to study the monetary transmission mechanism in these countries. Franta *et al.* (2014) use a TVP VAR to evaluate monetary policy transmission in the Czech Republic. Darvas (2013) extends this study, considering two other countries in addition to the Czech Republic, namely Hungary and Poland.

In motivational terms, the chapter most closely follows the work from D'Agostino *et al.* (2013) and analyses to what extent identifying the changes in the EECs' inflation and output growth dynamics helps to improve forecasting their macroeconomic variables. In order to implement my analysis, I use four time series, namely the consumer inflation, GDP growth, the short-term domestic interest rate and the real effective exchange rate. The number of time series is kept low to avoid that the number of parameters becomes excessively large relative to the available observations. It is worth noting that, in contrast to studies investigating large economies, such as the one undergone by D'Agostino *et al.* (2013) themselves, adding the real effective exchange rate variable is necessary since all chosen EECs are *de facto* small open economies.

The results of the analysis can be summarized as follows. First, I show that allowing for the parameters to vary improves the forecasting performance. Especially regarding Czech and Hungarian data, the time varying parameters vector autoregression (TVP-VAR) performs on average substantially better than the other models. Concerning Polish data, conversely, the vector autoregressive (VAR) model with fixed parameters performs overall slightly better. Second, the time varying parameters autoregressive (TVP-AR) model does particularly well in forecasting GDP growth. Third, as far as forecasts in larger horizons are involved, the threshold vector autoregressive (TAR) model is generally capable to predict troughs better than models systematically allowing for time varying parameters.

The chapter is organized as follows. In Section 3.1, I describe the different models and techniques that I use for my analysis. Section 3.2 provides an overview about the chosen priors, illustrates the data set and summarizes the empirical findings. Section 3.3 concludes, suggesting some potential future extensions and development.

3.1 The models

In my analysis, I compare the performance of several linear as well as non-linear models. Specifically, I use the time-varying parameter autoregressive (TVP-AR) model as a benchmark, and compare its results with those obtained by other non-linear models such as the threshold vector autoregressive (TAR) model and the time-varying parameter vector autoregressive (TVP-VAR) model. I also include in the comparison the results obtained by a vector autoregressive (VAR) model with fixed coefficients. In order to evaluate the relative performance of the models, I use the root mean squared error (RMSE) and the log-score for a horizon of eight forecasting periods.

Regarding the length of the autoregressive structure, the literature generally opts for two lags when dealing with quarterly data. This is in line with the work of Junicke and Merella (2017) and Maćkowiak (2006), who estimate their models using monthly data with six lags as an optimum. In my quarterly data, both the Akaike and the Schwartz criteria confirm that a VAR(2) estimation provides the best fit. However, for the non-linear model analysis I use a model with only one lag, in order to avoid too high a number of parameters relative to the limited available dataset, which might lead to a large estimation error.

In the early stages of the VAR literature, econometric models were usually estimated by using ordinary least squares (OLS) methods. Recently, Bayesian methods have attracted increased attention, because they are generally more flexible and precise than standard estimation approaches. Compared to the standard methodology, Bayesian estimations incorporate subjective beliefs or theoretical restrictions about the state of the coefficients. Bayesian methods were introduced by Zellner (1971), and have become quite popular in the last twenty years, in the wake of the enhanced computer capabilities and novel techniques such as the Gibbs sampling method were developed.[1] In this work, I follow the most recent literature and apply Bayesian estimation methods using Gibbs sampling to estimate the parameters of the autoregressive models listed above. The posterior distribution of the model is estimated also using Gibbs sampling, following Kim and Nelson (1999) and Koop (2003). Gibbs sampling is a convenient estimation method that enables to implement prior beliefs about the estimated parameters, and calculates the posterior distribution proportional to the product between the likelihood and the prior distribution, according to the Bayes theorem.

The Bayesian estimation combines a subjective prior together with sample information. It is based on Bayes theorem, which states that

$$\text{posterior distribution} \propto \text{likelihood} \times \text{prior distribution}$$

The likelihood function is taken from the OLS estimation of the data sample.[2]

[1] See Greene (2003), Chapter 18, for further details.

[2] The likelihood function for B and Σ, conditional on the data, is expressed by

$$F\left(Y \backslash vec\left(B\right), \Sigma\right) = \left(2\pi\right)^{-Tm/2} \left|\Sigma^{-1}\right|^{T/2} \exp\left[-\frac{1}{2}\left(y - vec\left(B\right)' X\right)' \Sigma^{-1} \left(y - vec\left(B\right)' X\right)\right].$$

For more details, see, *e.g.*, Hamilton (1994), Chapter 11.

Equivalently, it can be written as

$$G(b \backslash \Sigma, Y_t) \propto F(Y_t \backslash b, \Sigma) \times P(b, \Sigma)$$

where the posterior distribution $G(b \backslash \Sigma, Y_t)$ is proportional to the product of the prior distribution $P(b, \Sigma)$ and distribution of the sample, given by the likelihood function $F(Y_t \backslash b, \Sigma)$. The vector b represents the matrix of regressors B in vector form, and Σ the variance-covariance matrix. The prior density and the likelihood function are both very important for the correct estimation of the model and therefore it is necessary to give each a full specification. There exist several approaches to set the prior. Many authors use the Minnesota prior, developed by Litterman (1986), because of its simplicity. In this work I follow the more recent approaches in the literature, and use an independent normal inverse Wishart prior, which can be imposed by incorporating additional artificial data. This approach was first implemented by Bańbura *et al.* (2010) and it is more suitable for models with a limited number of observations as it is the case here.

3.1.1 Fixed coefficient VAR model

The recursive VAR model with fixed coefficients is the most general formulation within the VAR family. Consider T observations of m variables. Take a VAR(p) process, where p is the number of lags of the process with linear structure, to estimate the relationship among a set of endogenous variables as follows

$$Y_t = X_t B + \epsilon_t$$

with $X_t = (Y_{t-1}, ...Y_{t-p}, 1)'$, where Y_t is a $m \times 1$ vector of endogenous variables in period t. The intercept term C is a $m \times 1$ vector, which allows for the possibility of a nonzero $E[Y_t]$. B is the $m \times m$ matrix of regressors

$$B = (B_{1,}...B_p, C)$$

The residual ϵ_t is a Gaussian white noise with zero mean and variance-covariance matrix Σ, such that

$$\begin{aligned} E[\epsilon_t \epsilon_s'] &= \Sigma \text{ if } t = s \\ E[\epsilon_t \epsilon_s'] &= 0 \text{ if } t \neq s \end{aligned}$$

I can impose restriction on the contemporaneous relationships using a Cholesky decomposition in such a way that

$$\epsilon_t = A e_t$$

where A is a lower triangular matrix and it holds that

$$\begin{aligned} E[e_t e_s'] &= I_m \text{ if } t = s \\ E[e_t e_s'] &= 0 \text{ if } t \neq s \end{aligned}$$

and therefore

$$\Sigma = AA'$$

Note that, since each equation in the VAR has identical regressors, the model can be re-written as

$$y = (I_N \otimes X)\, b + V$$

where $y = vec\,(Y_t)$, $b = vec\,(B)$ and $V = vec\,(v_t)$. Assume that the prior for the VAR coefficients b is normally distributed and given by

$$p\,(b) \sim N\left(\widetilde{b}_0, H\right)$$

where $\widetilde{b}_0 = vec\,(B_0)$ is a $(m \times (m \times p + 1)) \times 1$ vector indicating the prior mean, while H is a is a $(m \times (m \times p + 1)) \times (m \times (m \times p + 1))$ matrix where the diagonal elements denote the variance of the prior. I discuss the ways of setting $\widetilde{b}_0$ and H in detail below.

Given the fact that the conjugate prior on B is normal distributed (see, *e.g.*, Blake and Mumtaz, 2012; and Kadiyala and Karlsson, 1997), it can be shown that the posterior distribution of the coefficients conditional on the variance-covariance matrix Σ is given by

$$G\left(b \backslash \Sigma, Y_t\right) \sim N\left(M^*, V^*\right)$$

where M^* and V^* are the mean and the variance of the normal distribution, respectively. As shown in Hamilton (1994, p.354) and in Kadiyala and Karlsson (1997), the mean and the variance of this normal distribution are given by the following expressions

$$\begin{aligned} M^* &= \left(H^{-1} + \Sigma^{-1} \otimes X_t' X_t\right)^{-1} \left(H^{-1} \widetilde{b}_0 + \Sigma^{-1} \otimes X_t' X_t \widehat{b}\right) \qquad (3.1) \\ V^* &= \left(H^{-1} + \Sigma^{-1} \otimes X_t' X_t\right)^{-1} \end{aligned}$$

where $\widehat{b}$ is a $(m \times (m \times p + 1)) \times 1$ vector indicating the OLS estimates of the VAR coefficients in a vectorized format, $\widehat{b} = vec\left(\left(X_t' X_t\right)^{-1} \left(X_t' Y_t\right)\right)$. Note that the conditional posterior mean M^* is a weighted average of the prior mean $\widetilde{b}_0$ and the OLS estimator, given by $X_t' Y_t$, weighted by the reciprocal of the corresponding variance-covariance matrices. The inverse variance is given by $\Sigma^{-1} \otimes X_t' X_t$ for $\widehat{b}$ is and by H^{-1} for$\widetilde{b}_0$. Note also that the smaller the values of matrix H elements, the higher the weight on the prior relative to the conditional posterior estimates. In the case where there are no beliefs about the prior, *i.e.*, the value of matrix H

elements are very large, then the posterior estimates are identical to the maximum likelihood estimator.

Following Zellner (1971), the conjugate prior for a positive definite VAR variance-covariance matrix Σ is an Inverse Wishart prior, with prior scale matrix $\overline{S}$ and prior degree of freedom α

$$p(\Sigma) \sim IW\left(\overline{S}, \alpha\right)$$

Given the fact that conjugate prior on B is normal distributed, it can be shown that the posterior distribution of the coefficients conditional on the variance-covariance matrix Σ is given by

$$G(b \backslash \Sigma, Y_t) \sim N(M^*, V^*)$$

Thus, the posterior for variance matrix Σ conditional on b is also Wishart distributed

$$G(\Sigma \backslash b, Y_t) \sim IW\left(\overline{\Sigma}, T + \alpha\right)$$

where T is the size of the sample and

$$\overline{\Sigma} = \overline{S} + (Y_t - X_t B)' (Y_t - X_t B)$$

where B is again reshaped into a $(m \times p + 1)$ by m matrix.

Creating dummy observations to incorporate prior

Returning to the issue of how to incorporate prior beliefs into the estimation of my VAR model, I follow Bańbura *et al.* (2010) and implement prior information by adding artificial data to the system. Informally speaking, this way of implementing prior involves generating artificial data from the model assumed under the prior, and mixing this with the actual data. The weight placed on the artificial data determines how tightly the prior is imposed.

Consider artificial data denoted Y_D and X_D such that

$$\begin{aligned} B_0 &= \left(X_D' X_D\right)^{-1} \left(X_D' Y_D\right) \\ S_0 &= \left(Y_D - X_D B_0\right)' \left(Y_D - X_D B_0\right) \end{aligned}$$

In other words a regression of Y_D and X_D gives the prior mean for the VAR coefficients, and the sum of squared residuals give the prior scale matrix for the error covariance matrix. The prior takes the normal inverse Wishart form

$$\begin{aligned} p\left(B \backslash \Sigma\right) &\sim N\left(\widetilde{b}_0, \Sigma \otimes \left(X_D' X_D\right)^{-1}\right) \\ p\left(\Sigma\right) &\sim IW\left(S, T_D - K\right) \end{aligned}$$

where again $\widetilde{b}_0 = vec\left(B_0\right)$, T_D is the length of the artificial data and K denotes the number of regressors in each equation. Given this artificial data, the conditional posterior distributions for the VAR parameters are given by

$$\begin{aligned} G(B \backslash \Sigma) &\sim N\left(b^*, \Sigma \otimes (X^{*\prime} X^*)^{-1}\right) \\ G(\Sigma \backslash B, Y_t) &\sim IW(S^*, T^*) \end{aligned}$$

with $Y^* = [Y_t; Y_D]$ a $X^* = [X_t; X_D]$. Thus, the dataset is constructed by artificial and actual data. T^* denotes the the number of rows in Y^*. Furthermore, it holds that $b^* = vec(B^*)$ with

$$B^* = (X^{*\prime} X^*)^{-1} (X^{*\prime} Y^*)$$

and

$$S^* = (Y^* - X^* B^*)' (Y^* - X^* B^*)$$

The artificial data Y_D and X_D are formed by four independent blocks as follows

$$Y_D=\begin{pmatrix} \frac{diag(\chi_1\sigma_1...\chi_m\sigma_m)}{\lambda} \\ 0_{((m\times(p-1))\times m)} \\ ... \\ diag\,(\sigma_1...\sigma_m) \\ ... \\ 0_{(1\times m)} \\ ... \\ \frac{diag(\chi_1\mu_1...\chi_m\mu_m)}{\tau} \end{pmatrix},\ X_D=\begin{pmatrix} \frac{J_p\otimes diag(\sigma_1...\sigma_m)}{\lambda} & 0_{(mp\times 1)} \\ ... & ... \\ 0_{(m\times mp)} & 0_{(m\times 1)} \\ ... & ... \\ 0_{(1\times mp)} & c \\ ... & ... \\ \frac{J_p\otimes diag(\chi_1\mu_1...\chi_m\mu_m)}{\tau} & 0_{(m\times 1)} \end{pmatrix}. \qquad (3.2)$$

The first block in each matrix imposes the prior beliefs on the autoregressive coefficients. The second block implements the prior for the variance-covariance matrix and the third block reflects the uninformative prior for the intercept. By adding artificial data in the last row, I incorporate the prior expressing the belief that the sum of the coefficients on lags of the dependent variable in each equation sum to 1, *i.e.*, that each variable has a unit root. The matrix J_p is given as $J_p = diag\,(1...p)$. As in Bańbura *et al.* (2010), the variance of the prior distribution is defined by hyperparameters that regulate the variation around the prior. The hyperparameter $\lambda > 0$ controls the overall tightness of the prior so that, as $\lambda \to 0$, the prior is implemented more tightly, whereas the larger the value of this parameter, the more the posterior approaches an ordinary least square (OLS) estimation of the VAR model. The hyperparameter τ controls for the degree of shrinkage. If τ is large, the prior is imposed loosely. I set $\lambda = 10$ and $\tau = 10\lambda$, implying that the prior on these data is not very informative. The parameter χ_i measures the persistence of variable i, and follows from the OLS estimation of an AR(1). Literally, it is a prior mean for the coefficient on the first lag of dependent variable i. The parameter μ_i is a sample mean of the variable i, and σ_i is a sample standard deviation of error terms. They can both be calculated as sample averages of the time series y_i from the OLS estimation. The matrix Y_D is the $(m\,(p+2)+1) \times m$ matrix and X_D is a $(m\,(p+2)+1) \times (mp+1)$ matrix adding $(m\,(p+2)+1)$ dummies to each time series. These artificial data are mixing with the actual data and the hyperparameters placed on them determine how tightly the prior is imposed. This approach also helps to alleviate the curse of dimensionality in the VAR model.

Gibbs Sampling

To carry out the Bayesian inference, I use a Gibbs sampling procedure, which is a posterior Markov chain Monte Carlo (MCMC) simulation mechanism. Gibbs sampling is a numerical method that uses a great many draws from a conditional distribution to approximate joint and marginal posterior distribution for B and Σ. In other words, the Gibbs sampler simulates the posterior distribution of the unknown parameters. The reason for using Gibbs sampling to calculate the marginal density is that analytical methods are unavailable.

The Gibbs algorithm iterates M times and produces draws for B and Σ. Each iteration requires sampling from the conditional posterior distribution, which after the burn-in draws eventually converges to the marginal distribution. The burn-in draws are the samples from the beginning of the chain. The first J draws are discarded to remove the influence of starting values. Once draws from the posterior distribution are obtained, I implement a structural analysis to ensure that the sign restrictions hold.

The Gibbs algorithm can be summarized as follows.

1. Set the priors for the coefficient matrix $p\left(vec\left(B\right)\right) \sim N\left(vec\left(B_0\right), H\right)$ and for the variance-covariance matrix $p\left(\Sigma\right) \sim IW\left(\bar{S}, \alpha\right)$ as described above, with starting values obtained from the OLS estimation.

2. Sample the conditional posterior distribution of B, the first coefficient of vector $vec\left(B_1\right)$, with variance V^* and mean M^* as given by equation (3.1).

3. Given $vec\left(B_1\right)$, draw variance-covariance matrix Σ_1 from the Inverse Wishart distribution.

4. Compute a matrix A, such that $AA' = \Sigma$ using a Cholesky decomposition.

5. Save matrix A to use it for further analysis.

6. Repeat 1-6 M times to obtain $vec\left(B_1\right), ..., vec\left(B_M\right)$, $\Sigma_1, ...\Sigma_M$, and discard the first J iterations. Use the remaining $M-J$ iterations to approximate the marginal posterior distribution, and the posterior mean and variance.

I set $M = 50000$ iterations, of which the first $J = 40000$ are discarded, thus keeping $M-J$ draws to use for further inference. That is, the last 10000 values of B and Σ of the full set iterations are used to form the empirical distribution of these parameters. Note that these draws of these model parameters are also used to calculate the forecast density

$$G\left(Y_{t+k} \backslash Y_t\right) = \int G\left(Y_{t+k} \backslash Y_t, \Gamma\right) \times G\left(\Gamma \backslash Y_t\right) d\Gamma$$

where $k = 1, 2, ...8$ is the forecasting horizon, and $\Gamma = \{B, \Sigma\}$. The forecast density can be easily obtained by simulating Y_t variables k periods forward, just using the last $M-J$ values of the Gibbs sampling for B and Σ.

Forecast

Let $y^T = [y_1'...y_T']'$ be the time series data up to time T, and let $y^{T+1,T+h} = \left[y_{T+1}'...y_{T+h}'\right]'$ be the time series to forecast over the horizon $T+h$. Forecasts are obtained using the posterior forecast density conditional on the currently observed data

$$p\left(y^{T+1,T+h} | y^T\right)$$

Hence, forecasts are obtained by estimating the above models iterating for $h = 8$ periods ahead. Considering a VAR model of the form

$$Y_t = c + BY_{t-1} + \varepsilon_t$$

it follows that forecast for $h = 1$ obtained using this model is optimal if it holds

$$\hat{Y}_{t+1} = E(Y_{t+1}|I_t) = c + BY_t$$

where I_t is information set at time t, and c is the time trend (constant). Analogously, for the second period ahead, it holds

$$\hat{Y}_{t+2} = c + B\hat{Y}_{t+1} = c + B(c + BY_t) = c + Bc + B^2 Y_t$$

where B^2 denotes the product BB'. More generally, we can write a forecast h periods ahead such as

$$\hat{Y}_{t+h} = c + Bc + B^2 c + ... + B^{h-1} c + B^h Y_t$$

Thus, for a forecast with an infinite horizon it holds

$$\hat{Y}_{t+\infty} = (\Sigma_{i=1}^{\infty} B)c = (I - B)^{-1} c = E[Y_t]$$

3.1.2 Threshold VAR model

The Threshold vector autoregressive (TAR) model belongs to a group of non-linear models that allow for the parameters to vary in different regimes. It is formally defined as

$$Y_t = \left[c_1 + X_t B_1 + \Sigma_1^{1/2} e_t\right] S_t + \left[c_2 + X_t B_2 + \Sigma_2^{1/2} e_t\right] (1 - S_t) \tag{3.3}$$

where the residual e_t is a Gaussian white noise with zero mean and identity variance-covariance matrix.

The state variable S_t is a dummy variable taking the values 1 or 0 according to the following rule

$$\begin{aligned} S_t &= 1 \text{ if } Z_{t-d} \leq Z^* \\ & \quad 0 \text{ otherwise} \end{aligned} \tag{3.4}$$

Thus, the model allows for the possibility of two regimes, where the regime is determined by the level of a threshold variable Z_{t-d} relative to an unobserved threshold level Z^*. The two sets of parameters $\{c_j, B_j, \Sigma_j\}$, with $j = 1, 2$, can be regarded as the reduced-form counterparts of two sets of first-order conditions associated to a structural model with an occasionally binding borrowing constraint, and corresponding respectively to the states where the constraint does or does not bind.

The switching mechanism of the TAR methodology is intuitive and simple. Unlike time-varying parameter models, the time-variation in the parameters is linked explicitly to a threshold variable. As the threshold variable, I choose economic growth. As a result, the parameters are allowed to be different in expansions and recessions. In the application, the threshold variable is assumed to be lagged d^{th} time, where the delay d is assumed to be an unknown parameter. I assume a flat prior on the delay d, and limit its values between 1 and 4. This is in line with the recent literature on threshold VARs: for instance, Alessandri and Mumtaz (2017) opt for an upper bound of 12 lags in a monthly dataset.

In order to estimate the TAR, I use Gibbs sample algorithm similar to the one for employed for the VAR, adjusted to deal with some specificities of the TAR model. In particular, I impose a natural conjugate prior on the VAR parameters in the two regimes. The prior tightness is set in an identical fashion to the previous VAR case. We assume a normal prior for $Z^* \sim N\left(\bar{Z}, \bar{V}\right)$, where $\bar{Z} = 1/T \sum_{i=1}^{T} Z_t$ and $\bar{V} = 10$. This represents a fairly loose prior. Given an initial value for Z^* and d, the conditional posterior for the TAR parameters in the two regimes is normal Wishart distributed and, analogously to that found in Section 3.1.1, given by

$$G\left(B \backslash \Sigma\right) \sim N\left(b^*, \Sigma \otimes \left(X^{*\prime} X^*\right)^{-1}\right)$$

$$G(\Sigma \backslash B, Y_t) \sim IW(S^*, T^*)$$

Since the posterior distribution of Z^* conditional to the TAR parameter vector is unknown, following Chen and Lee (1995), Chen (1998) and Lopes and Salazar (2006), I employ both the Gibbs sampler and the Metropolis-Hasting samplers to derive the full posterior distribution of the entire estimated parameters vector.[3]

For a given draw for the TAR parameters and a value for d, a random walk Metropolis-Hastings step can be employed to sample Z^*. We draw a candidate value as $Z^*_{new} = Z^*_{old} + \Psi^{1/2}e$, with $e \sim N(0,1)$. The acceptance probability is given by

$$\frac{f(Y_t \backslash Z^*_{new}, \Xi)}{f(Y_t \backslash Z^*_{old}, \Xi)}$$

where $f(.)$ denotes the posterior density, and Ξ represents all other parameters in the model. We choose the scaling factor Ψ to ensure that the acceptance rate remains between 20% and 40%. Chen and Lee (1995) show that the conditional posterior for d is a multinomial distribution with probability

$$\frac{L(Y_t \backslash d, \Xi)}{\sum_{d=1}^{m} L(Y_t \backslash d, \Xi)}$$

where $L(.)$ denotes the likelihood function and m is the maximum number of lags

[3]The Metropolis-Hasting alghoritm is more general than Gibbs alghorithm, and it used when the conditional posterior distribution is not available in closed form.

set in the prior. Again, we assume a TAR with two lags and use the Gibbs sampling procedure with $M = 50000$ iterations, discarding the first $J = 40000$ iterations. The $M - J$ draws are kept and used for further inference.

The forecast density for the TAR is defined as

$$G\left(Y_{t+k} \backslash Y_t\right) = \int G\left(Y_{t+k} \backslash Y_t, \Gamma\right) \times G\left(\Gamma \backslash Y_t\right) d\Gamma$$

where

$$\Gamma = \{B_1, \Sigma_1 B_2, \Sigma_2, Z^*, d\}$$

To proceed with the forecast, we simulate Y_{t+k} variables by iterating equations 3.3 and 3.4 k periods in the future.[4]

3.1.3 Models with time-varying parameters

Time varying parameter models with stochastic volatility gained recently particular attention in the literature. Contrary to linear models, they enable to let coefficients and variance vary with time. First, we consider a time varying parameter AR model with stochastic volatility (TVP-AR). Following Blake and Mumtaz

[4]Further technical details on this procedure can be found in Barnett *et al.* (2014, Appendix F) and in Alessandri and Mumtaz (2017, p.18). For a code running the algorithm, see http://cremfi.econ.qmul.ac.uk/efp/info.php.

(2012, p.126), the model takes the form

$$y_t = c_t + b_t y_{t-1} + \epsilon_t \sqrt{\exp\left(\ln(h_t)\right)}$$

where y_t is the endogenous variable, $B \equiv (c, b)$ is a vector of parameters and $\epsilon_t \sqrt{\exp\left(\ln(h_t)\right)}$ is a stochastic error term. The model reflects two types of volatility. One the one hand, time-varying coefficients B, whose variation can be due to changes in the structural dynamics, follow the random walk process

$$B_t = B_{t-1} + \varepsilon_t$$

with a Gaussian white noise $\varepsilon_t \sim N\left(0, Q\right)$.

Second, time-varying variance of the error term h_t follows itself a random walk process, formally

$$\ln(h_t) = \ln(h_{t-1}) + v_t$$

with $v_t \sim N\left(0, g\right)$.

The prior for matrix for Q is Inverse Wishart distributed, and given by

$$p\left(Q\right) \sim IW\left(\overline{Q}, \alpha\right)$$

where $\overline{Q}$ is a prior scale matrix and α is a prior degree of freedom. The prior scale matrix is

$$\overline{Q} = k \times Q_{OLS} \times T_0$$

where T_0 is the length of the training sample. The Q_{OLS} is the variance-covariance matrix of B obtained via ordinary least square estimation using the training sample. Furthermore, k is a scaling factor set to a small number. (I assume $k = e^{-5}$ to reflect that the training sample is quite short. As a consequence, the results following from the estimation may be relatively imprecise).

The prior for variance g is a inverse Gamma distributed[5]

$$p(g) \sim IG(g_0, v_0)$$

and all the starting values are set using the OLS estimation of the training sample. We set the training sample to be formed of $T_0 = 30$ quarters.

Analogously, following once again Blake and Mumtaz (2012, p.92), I assume that the time-varying parameters VAR with stochastic volatility can be described as

$$Y_t = X_t B_t + \epsilon_t \tag{3.5}$$

where Y_t is a $m \times 1$ vector of endogenous variables in period t and

[5] Inverse Gamma distribution is the univariate analog of the inverse Whishart. In other words, one can think of the inverse Wishart distribution as a multivariate version of the inverse Gamma distribution.

$B_t = (B_{t,1}, ..., B_{t,p}, C_t)$ is a matrix of the time varying parameters. We assume that the parameter matrix B_t follows the random walk process

$$B_t = B_{t-1} + e_t$$

with variance-covariance matrix

$$Var\left(e_t\right) = Q$$

The residual ϵ_t is a Gaussian white noise with zero mean and variance-covariance matrix Σ_t. As in Cogley and Sargent (2005) and Barnett *et al.* (2014, p.14), without loss of the generality, we can decompose the variance-covariance matrix using the following structure

$$\Sigma_t = A_t^{-1} H_t A_t^{-1\prime}$$

where A_t is a lower triangular matrix and H_t is a diagonal matrix, formally

$$A_t = \begin{pmatrix} 1 & 0 & 0 \\ a_{12,t} & 1 & 0 \\ a_{13,t} & a_{23,t} & 1 \end{pmatrix} \qquad H_t = \begin{pmatrix} h_{1,t} & 0 & 0 \\ 0 & h_{2,t} & 0 \\ 0 & 0 & h_{3,t} \end{pmatrix}$$

Following Primiceri (2005), we postulate the non-zero and non-one elements of the matrix A_t to evolve as driftless random walks,

$$\begin{aligned} a_{ij,t} &= a_{ij,t-1} + V_t \\ Var\left(V_t\right) &= D \end{aligned}$$

while $h_{i,t}$ follows the random work

$$\begin{aligned} \ln h_{i,t} &= \ln h_{i,t-1} + z_{i,t}, \\ Var\left(z_{i,t}\right) &= g_i. \end{aligned}$$

As a result, it holds

$$A_t v_t = \varepsilon_t$$

with $Var\left(\varepsilon_t\right) = H_t$. Thus, this model has two sets of time varying coefficients, β and $a_{ij,t}$, and a time varying stochastic volatility, whose value is based on the diagonal elements $h_{i,t}$.

Following Blake and Mumtaz (2012), I can describe the sampling algorithm for the TVP-VAR model as follows.

1. Set the priors and starting values. For the covariance matrix Q, the starting values are set using OLS estimation, $p\left(Q\right) \sim IW\left(Q_0, T_0\right)$. Note that this prior is quite critical, as it influences the amount of time-variation allowed

for in the VAR model. In other words, a large value for the scale matrix Q_0 would allow for larger fluctuations in β_t. This prior is typically set using a training sample. The first T_0 observations of the sample are used to estimate a standard fixed coefficient VAR via OLS such that

$$\beta_0 = \left(X'_{0t}X_{0t}\right)^{-1}\left(X'_{0t}Y_{0t}\right)$$

where the subscript 0 denotes the fact that this is the training sample. Set the prior for D and the initial values for $a_{ij,t}$.

2. Conditional on A_t, H_t and Q, draw β_t.

3. Using the draw for β_t, calculate the residuals of the transition equation $\beta_t - \beta_{t-1} = e_t$. Sample Q from the Inverse Wishart distribution using the scale matrix $e_t e'_t + Q_0$ and the number of degrees of freedom $T + T_0$.

4. Draw $a_{ij,t}$ elements of the matrix A_t, conditional on β_t, H and D.

5. Calculate the residuals for V_t, conditional on the draws for $a_{ij,t}$. Draw D from the Inverse Wishart distribution.

6. Using the draw of A_t from step 4, calculate $\varepsilon_t = A_t v_t$. Draw $h_{i,t}$.

7. Conditional on the draw $h_{i,t}$, draw g_i.

8. Repeat steps 2 to 7 M times.

The prior for Q is set using a training sample of $T_0 = 30$ quarters. In particular, let Q_{OLS} denote the ordinary least square estimate using the training sample. Then the prior distribution for Q is assumed to be inverse Wishart distributed with a scale matrix $\bar{Q} = Q_{OLS} \times T_0 \times k$.

Because our dataset is quite short, we choose a model with one lag for the estimation of the models allowing for time variation. Assuming more lags would require an estimation of large number of parameters, which is more difficult using a TVP-VAR model. The models are estimated using a Gibbs sampling algorithm. The priors distributions and conditional posteriors are implemented via dummy observations, as described in previous section. The number of iterations is set by $M = 50000$, from which first $J = 40000$ iterations are discarded, and the remaining $M - J$ draws uses for further inference.

The forecasting procedure is similar to those for linear VAR model. A technical issue arises when multi-step expectations are generated, because it is necessary to evaluate the future paths of drifting parameters. I follow the literature and treat those parameters as if they had remained constant at the current level. As a consequence, forecasts at time $t + h$ are computed iteratively as shown in D'Agostino *et al.* (2013).

3.1.4 Forecasting evaluation

For my analysis, I set $Y_t = (\pi_t, y_t, i_t, reer)'$, where π_t is the inflation rate, y_t is output growth rate, i_t is the interest rate and $reer_t$ the real effective exchange rate. All models are estimated recursively over an expanding data window. To compare the average forecasting performance of the models, we choose two commonly used tools, the root mean squared error (RMSE) and the log predictive density scores described below.

Root Mean Squared Error

The root mean square error (RMSE) measures the quadratic differences between values predicted by a model and the values actually observed. Formally, RMSE is defined by the expression

$$RMSE = \sqrt{\sum_{t=T+1}^{T+h} \frac{(\mathbf{v}_{t+h})^2}{h}}$$

with forecast horizon $T+1, T+2, ..., T+h$. The expression $\mathbf{v}_{t+h}$ denotes the residual, given by the difference between forecasted and actual data. The h-step ahead forecast error is:

$$\begin{aligned}
\mathbf{v}_{t+h} &= Y_{t+h} - E(Y_{t+h}|I_t) \\
&= B^{h-1}\epsilon_{t+1} + ... + B\epsilon_{t+h-1} + \epsilon_{t+h} = \sum\nolimits_{j=1}^{h} B_{t+j}^{h-j}\epsilon
\end{aligned}$$

The variance of the h-step ahead forecast error is:

$$Var(\mathbf{v}_{t+h}) = B^{h-1}\Sigma B^{h-1\prime} + ... + B\Sigma B' + \Sigma$$

where Σ is the variance (matrix) of the error term ϵ_t.

In order to assess the performance of the different forecasting models, I compute the RMSE for each model and compare them across models. It follows that the lower the RMSE, the better the forecast performance of the model. Note that the forecast errors are correlated, and the correlation is measured by

$$\begin{aligned} Cov(\mathbf{v}_{t+2}, \mathbf{v}_{t+1}) &= Cov(B\epsilon_{t+1} + \epsilon_{t+2}, \epsilon_{t+1}) = B\Sigma \\ Cov(\mathbf{v}_{t+3}, \mathbf{v}_{t+1}) &= Cov(B^2_{t+1}\epsilon + B\epsilon_{t+2} + \epsilon_{t+3}, \epsilon_{t+1}) = B^2\Sigma \\ Cov(\mathbf{v}_{t+h}, \mathbf{v}_{t+1}) &= Cov(B^{h-j}_{t+j}\epsilon + ..., \epsilon_{t+1,}) = B^{h-j}\Sigma \end{aligned}$$

Log Predictive Density Score

Log Predictive Density Score is a measure of the forecast accuracy of the density. The measure is based on the posterior predictive distribution. In turn, this distribution is unobserved conditional on observed data and is derived by computing maximum likelihood estimate of the parameters given the observed data. These parameters are plugged into the distribution function of the new observations.

To assess the models' forecasting performance, we compare the log predictive density score of each model with the benchmark (as discussed above, the TVP-AR(1) process) in such a way that

$$L = LS_t^A - LS_t^{Benchmark}$$

where

$$LS_t^i = p\left(x_t^0 | Y_{t-1}, i\right)$$

is the predictive density generated by model i for x_t evaluated for the actual observation x_t^0. The model under consideration performs better than the benchmark at some forecast horizon $t+h$ if $L > 0$.

3.2 Empirical results

3.2.1 Data

The dataset consists of data with a quarterly frequency available on Macrobond and Fred. I use four time series, namely, inflation, GDP growth rate, short term interest rate and real effective exchange rate. Inflation is measured in terms of quarter on quarter growth rate of the seasonally adjusted consumer price index. GDP growth rate is measured as quarter on quarter change of the seasonally adjusted

GDP time series. The short term interest rate is measured by the three-month interbank interest rate. The real effective exchange rate (REER) is the weighted average of a country's currency relative to an index of other major currencies, adjusted for the effects of inflation. The weights are determined by comparing the relative trade balance of a country's currency against each country within the index.

The time span of the dataset is from 1997 Q1 to 2016 Q2 for the Czech Republic, from 1998 Q1 to 2016 Q2 for Hungary, and from 1995 Q1 to 2016 Q2 for Poland. GDP growth time series for Poland is the combination of two sources: the period 1995-2002 is covered by the FRED database, whereas the period 2014-2016 data are taken from Macrobond.

3.2.2 Forecasting results

Tables 3.1-3.3 report the estimated RMSEs for each model and country under consideration at the one, two, four and eight quarter forecasting horizons. The results are evaluated over the full sample for all three countries. In each table, the TVP-AR(1) model is the benchmark, and the RMSEs are reported in relative terms to the performance of the benchmark model. Thus, if the RMSE is above one, the forecasting model perform worse than the TVP-AR(1) model and, as such, the benchmark model is preferable. It is easy to notice that both the TAR model and the VAR model perform worse than the benchmark model. The results are further analyzed using the log-scores results, illustrated in Tables 3.4-3.6.

Inflation

The forecasting performance for inflation in the Czech Republic show mixed results across the different models. In terms of RMSE, as shown in Table 3.1, the standard VAR model with fixed parameters delivers the best results for inflation the first half a year. However, in the forecasting horizons of one to two years, the time varying version performs slightly better. The TAR model's performance is very close to the TVP-AR model at all the horizons, while both versions of the VAR model consistently outperform the benchmark. Considering the log-scores (the relevant results are reported in Table 3.4), the VAR model performs in average best at all the horizons, except for the four-quarter case (for which the best performer is the TVP-VAR model, though the VAR model is a close second). Overall, the VAR model can be considered the most suitable one to predict the Czech inflation.

The RMSE for inflation in Hungary, reported in Table 3.2, suggest that the TVP-VAR model clearly outperforms all other estimation models. The TAR model and the VAR model follow a very close pattern over the different horizons. They are both outperformed by the benchmark model at all horizon but the longest one. The log predictive density scores, illustrated in Table 3.5, confirm that the TVP-VAR

Inflation	1	2	4	8
VAR	0.6795	0.7784	0.8803	0.9330
TAR	1.0058	1.0063	1.0067	1.0066
TVP-VAR	0.8130	0.8260	0.8793	0.9153

GDP	1	2	4	8
VAR	0.8402	1.0812	1.4161	1.6288
TAR	0.9572	1.1048	1.3801	1.5253
TVP-VAR	0.7954	1.0237	1.3641	1.6436

Interest rate	1	2	4	8
VAR	0.3452	0.4473	0.6732	1.0655
TAR	1.4773	1.3993	1.3405	1.2762
TVP-VAR	0.2141	0.2938	0.4144	6.5997

REER	1	2	4	8
VAR	0.0447	0.0584	0.0885	0.1321
TAR	9.7466	9.0349	8.8218	9.2826
TVP-VAR	0.0442	0.0566	0.0804	0.2293

Table 3.1. RMSE prediction statistics for the Czech Republic

Inflation	1	2	4	8
VAR	1.2939	1.1100	1.0618	0.9448
TAR	1.3034	1.1168	1.0678	0.9493
TVP-VAR	0.9314	0.8753	0.8976	0.8467

GDP	1	2	4	8
VAR	1.5039	1.5910	1.6401	1.5626
TAR	1.5049	1.5915	1.6397	1.5623
TVP-VAR	1.4561	1.5775	1.6910	1.6406

Interest rate	1	2	4	8
VAR	4.0700	3.6850	3.3505	3.0856
TAR	4.1123	3.7241	3.3876	3.1224
TVP-VAR	0.4646	0.6369	0.8977	1.2037

REER	1	2	4	8
VAR	0.0548	0.0611	0.0699	0.0855
TAR	7.2714	6.2554	5.7943	5.6388
TVP-VAR	0.0530	0.0583	0.0662	0.0740

Table 3.2. RMSE prediction statistics for Hungary

Inflation	1	2	4	8
VAR	0.9566	0.9315	0.9245	0.9154
TAR	0.8855	0.9025	0.9052	0.8914
TVP-VAR	0.8862	0.9022	0.9049	0.8907

GDP	1	2	4	8
VAR	0.4866	0.4679	0.4659	0.4719
TAR	0.8083	0.8003	0.8012	0.7958
TVP-VAR	0.8005	0.7930	0.7945	0.7893

Interest rate	1	2	4	8
VAR	0.2444	0.3409	0.5185	0.8554
TAR	4.2802	3.8882	3.6557	3.5680
TVP-VAR	4.2377	3.8497	3.6187	3.5340

REER	1	2	4	8
VAR	0.0567	0.0715	0.0878	0.1047
TAR	7.8690	7.1996	6.8753	6.8877
TVP-VAR	0.0614	0.0773	0.0929	0.1054

Table 3.3. RMSE prediction statistics for Poland

Inflation	1	2	4	8
VAR	0.3659	0.2307	0.1451	0.3907
TAR	-0.5798	-0.5643	-0.6162	-0.4993
TVP-VAR	0.2194	0.2041	0.1662	0.3135

GDP	1	2	4	8
VAR	1.4158	-2.9415	-0.93	-3.0158
TAR	1.4394	-0.0197	-0.5246	-0.8356
TVP-VAR	1.7680	-0.0476	-0.7201	-0.9597

Interest rate	1	2	4	8
VAR	2.1374	0.9661	0.4725	0.7811
TAR	0.4083	-0.2794	-0.274	0.3875
TVP-VAR	2.5333	1.3135	0.8627	1.2238

REER	1	2	4	8
VAR	3.4664	1.9128	2.223	2.0360
TAR	-2.4926	-2.5264	-2.6658	-2.6556
TVP-VAR	3.5055	2.9972	2.4391	2.2742

Table 3.4. Log-scores prediction statistics for the Czech Republic

Inflation	1	2	4	8
VAR	-0.6031	-0.4001	-0.1225	0.2479
TAR	-0.6349	-0.3752	-0.3044	0.0724
TVP-VAR	0.0685	0.1354	0.1733	0.5524

GDP	1	2	4	8
VAR	-0.0447	-0.4582	-0.0711	0.4680
TAR	-0.0330	-0.0174	0.3290	0.5323
TVP-VAR	-0.1492	-0.3100	0.2928	0.5312

Interest rate	1	2	4	8
VAR	-3.5934	-2.9504	-2.2538	-1.1270
TAR	-1.4291	-1.3915	-1.1533	-0.6191
TVP-VAR	0.8191	0.2640	-0.1261	-0.2395

REER	1	2	4	8
VAR	3.2489	2.8073	2.6425	2.3634
TAR	-2.1548	-2.1558	-2.1529	-2.2039
TVP-VAR	3.3135	2.8881	2.7367	2.4295

Table 3.5. Log-scores prediction statistics for Hungary

Inflation	1	2	4	8
VAR	0.1316	0.1673	0.1744	0.1791
TAR	-0.5259	-0.5045	-0.4722	-0.4852
TVP-VAR	0.1471	0.1682	0.0851	-0.1617

GDP	1	2	4	8
VAR	0.9363	1.0208	1.0092	0.9901
TAR	0.2779	0.3327	0.3488	0.3403
TVP-VAR	-0.2096	0.4011	0.5733	0.4005

Interest rate	1	2	4	8
VAR	0.6507	0.0974	-0.4045	-0.6195
TAR	-1.1483	-1.1987	-1.1953	-0.9679
TVP-VAR	-3.7860	-3.8464	-2.5034	-1.2825

REER	1	2	4	8
VAR	2.7022	1.5678	1.935	1.8223
TAR	-2.6408	-2.7036	-2.7129	-2.7135
TVP-VAR	2.7205	2.2059	2.0509	2.0833

Table 3.6. Log-scores prediction statistics for Poland

performs best. The scores also confirm the similar patterns followed by the TAR model and the VAR model, which outperform the benchmark only at the longest horizon. Overall, the TVP-VAR model can be considered the most suitable one to predict the Hungarian inflation.

About Poland, considering the RMSE for inflation reported in Table 3.3, it is straightforward to notice that all models outperform the benchmark. The TAR model and TVP-VAR model show greater prediction accuracy than the VAR at all horizons. The performance of the two models is very similar, with the TAR model doing slightly better at the shortest horizon, and the TVP-VAR at the remaining, longest forecasted horizons. The log predictive density scores, illustrated in Table 3.6, deliver quite substantially different results. The benchmark outperforms the TAR model at all horizons, and also the TVP-VAR at the longest horizon. At the first two horizons, however, the TVP VAR is the most accurate predictor, whereas the VAR with fixed parameters performs substantially better at longer horizons. Overall, the VAR model can be considered the most suitable one to predict the Polish inflation, with the TVP-VAR model close second.

Output growth

For the Czech Republic, the RMSE reported in Table 3.1 suggest that the benchmark model outperforms all the other models at the three longest horizons in forecasting output growth. At the one-quarter horizon, the best performer is the TVP VAR model. Furthermore, also the TAR model and the VAR model show better performance than the benchmark. The results are also confirmed by log predictive density scores, illustrated in Table 3.4. Overall, the benchmark TVP-AR model can be considered the most suitable one to predict Czech output growth, though it performs quite poorly at a one-quarter horizon.

The RMSE results for economic growth in Hungary, reported in Table 3.2, suggest that the benchmark outperforms any other model at all forecasting horizons. The log predictive density scores, illustrated in Table 3.5, confirm this result on at the two shortest horizons. While the benchmark still outperforms the VAR at the four-quarter horizon, the other two models provide more accurate predictions, with the TAR model performing best. At the longest horizon, all models outperform the benchmark, and the TAR model still performs best, though the TVP-VAR model is close second. Overall, the benchmark TVP-AR model can be considered the most suitable one to predict Hungarian output growth, though the TAR model should be taken into account for the longer horizons.

Conversely, for Poland all models perform significantly better than the benchmark model according to the RMSEs, reported in Table 3.3. The VAR model with fixed coefficients clearly outperforms the alternative models, while the TAR model and the TVP-VAR model show similar predictive accuracy. The log predictive density scores, illustrated in Table 3.6, largely confirm these results, though it should be noted that, at the one-quarter forecasting horizon, the TVP-VAR is outperformed by the benchmark. Overall, the VAR model with fixed coefficients can clearly be considered the most suitable one to predict Polish output growth.

Interest rate and real effective exchange rate

The RMSE prediction statistics, reported in Tables 3.1-3.2, for the interest rate and the real effective exchange rate (REER) suggest clearly that the TVP-VAR model is the best performer for both the Czech Republic and Hungary up to the four-quarter forecasting horizon. However, at the eight-quarter horizon, for the interest rate the TVP-VAR is clearly outperformed by the benchmark model. Also, considering Czech REER at the longest horizon, the TVP-VAR model is also outperformed by the VAR model, while it still provide the more accurate prediction on Hungarian data. The log predictive density scores, illustrated in Tables 3.4-3.5, confirm that the TVP-VAR is the best performer for both the interest rate and the REER, as far as Czech and Hungarian data are concerned.

In light of the RMSE result, reported in Table 3.3, the VAR model with fixed parameters clearly performs best on Polish data with regard to both the interest rate and the REER. Concerning the interest rate, the performance of the TAR model can be compared with that of the TVP-VAR model, and both models are largely outperformed by the benchmark. In terms of REER predictions, the TVP-VAR model does substantially better than the benchmark, while the TAR model offer a strikingly poor performance. The log predictive density scores, illustrated in Table 3.6, suggest that the interest rate is best predicted by VAR the model at the two shorter horizons, and by the benchmark model at the longer ones. The REER is best predicted by the TVP-VAR model, with the VAR model close second.

Forecasting performance at a two-year horizon

Generally, the results of my analysis suggest that models allowing for time variations perform better than the other models. Thus, allowing for time-varying coefficients and error covariance matrix does lead to an improvement in the forecasting performance. However, it is interesting to see how the models are capable to predict the development of the key variables at long forecasting horizons. In Figures 1-6 I examine the evolution of the RMSEs for inflation and GDP growth for each country at the eight-quarter horizon. Since the data window expands as we move towards the more recent data point, the figures can be interpreted as a graphical analysis of how a given model performance evolves when the number of data upon which the estimation is performed increases.

Figure 3.1 shows that, for the Czech Republic, the performance of the TVP-AR model for the longest forecasting horizon improves as the numbers of data

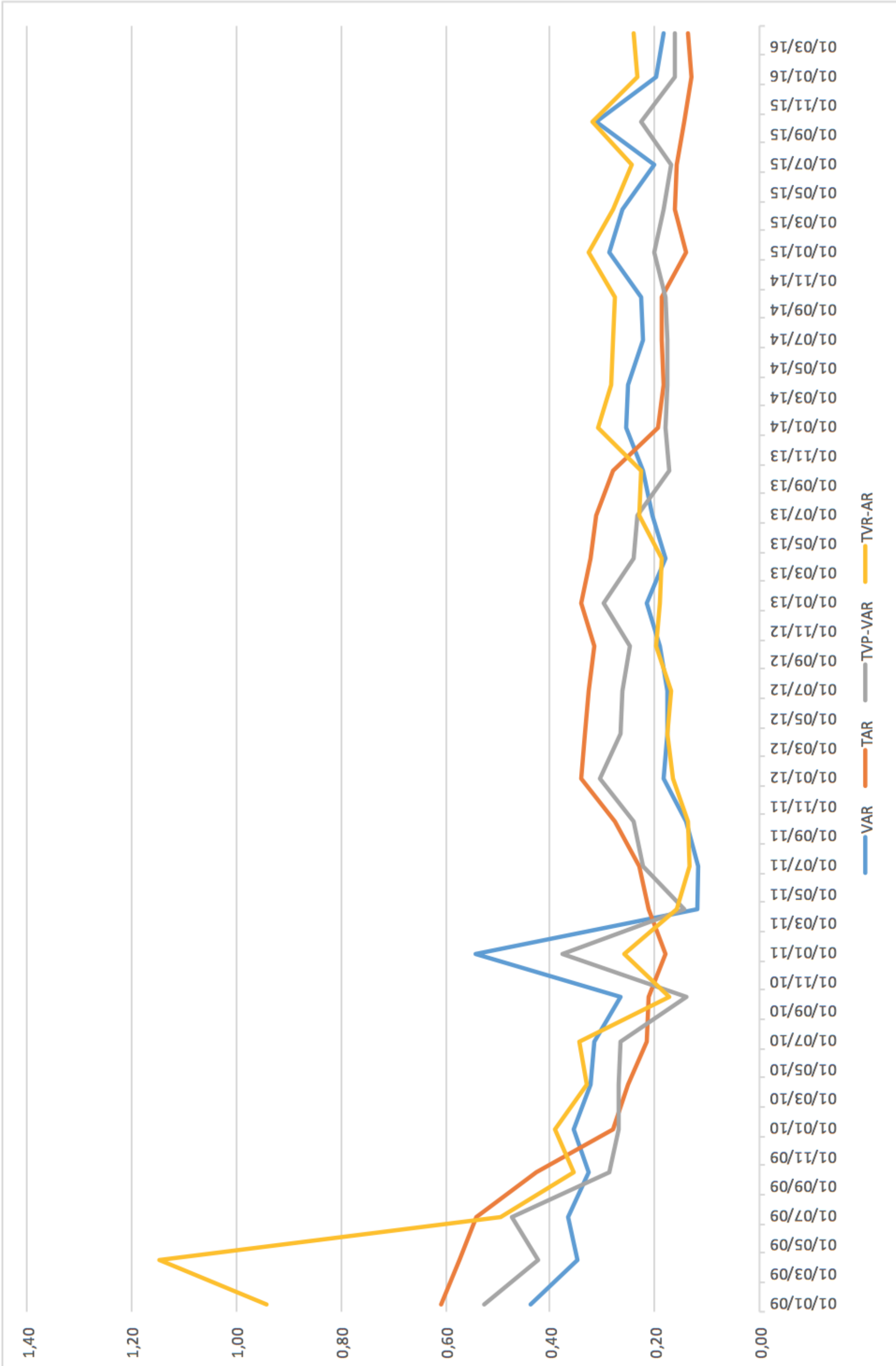

Figure 3.1. RMSE prediction statistics at the eight-quarter horizon for Czech inflation

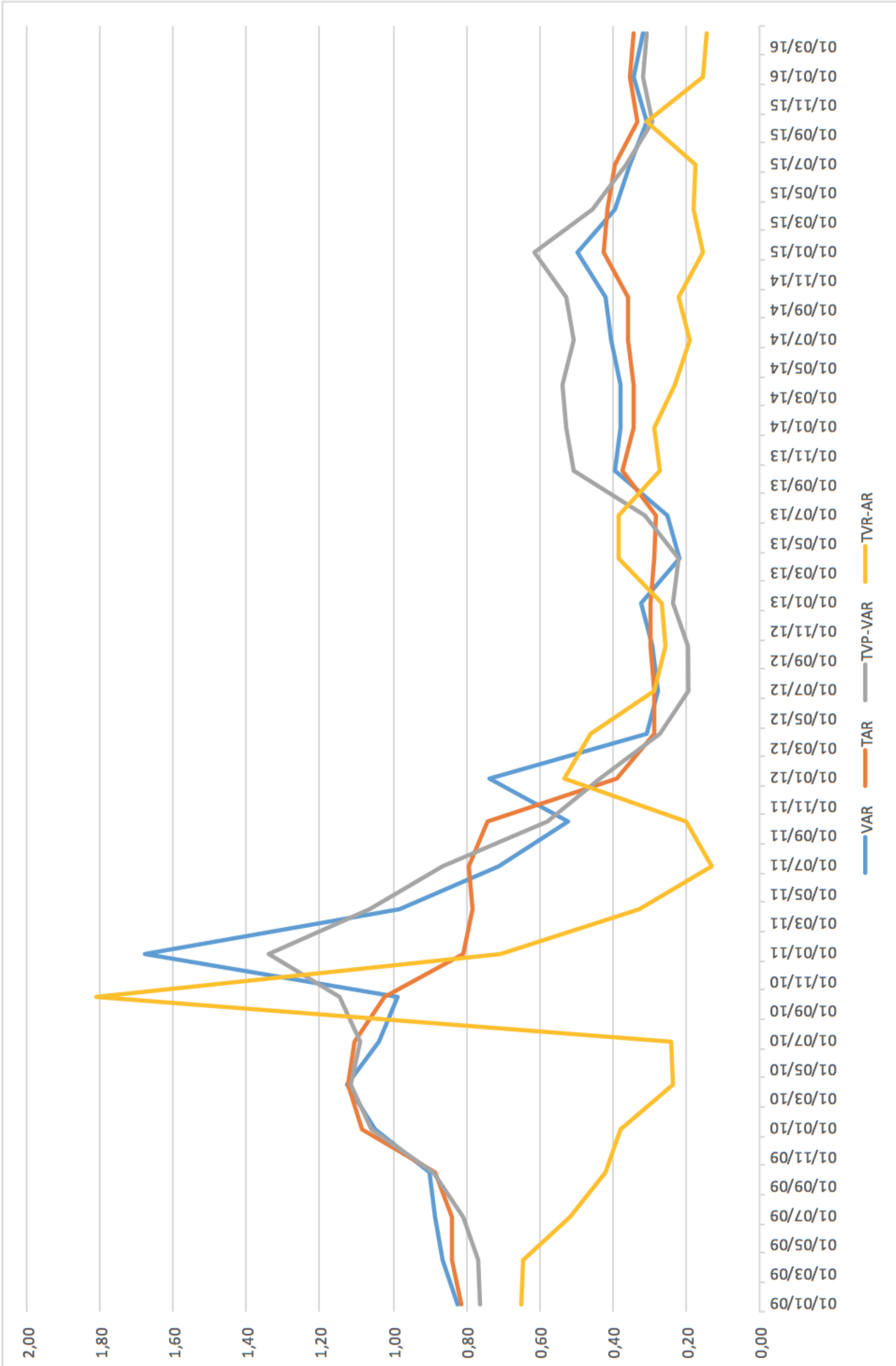

Figure 3.2. RMSE prediction statistics at the eight-quarter horizon for Czech output growth

involved in the estimation rises. As a result, at the first date of the time interval considered in the graph, the forecasting performance is poor relative to those of the other models. However, forecast accuracy recovers relatively quickly, and the TVP-AR model tends to outperform its competitor in time of 'normal' inflation. From November 2013, when the Czech National Bank started to fear deflation, the model looses again its predictive power, and is outperformed by both the TVP-VAR model and the TAR model. With regard to output growth, whose RMSEs are illustrated in Figure 3.2, the TAR model performs best in the times of crisis that characterize the years 2010 and 2011. Although the TVP-AR model is among the models unable to predict the crisis, it offers the highest predictive accuracy for the period after 2013.

For Hungarian inflation, whose RMSE are represented in Figure 3.3, the TAR model performs badly at the eight-quarter horizon in the times when inflation is higher. After 2013, when CPI inflation decreased, the TAR model is able to offer the better predictions than the other models. On average, the TVP-VAR model performs best, which can be seen especially during the period when the inflation was in line to the inflation target of the Hungarian central bank. About Hungarian output growth, whose RMSEs are illustrated in Figure 3.4, the group of the VAR models is clearly outperformed by the TVP-AR model almost for the entire duration of the considered period. Only for a short interval, in 2012, when

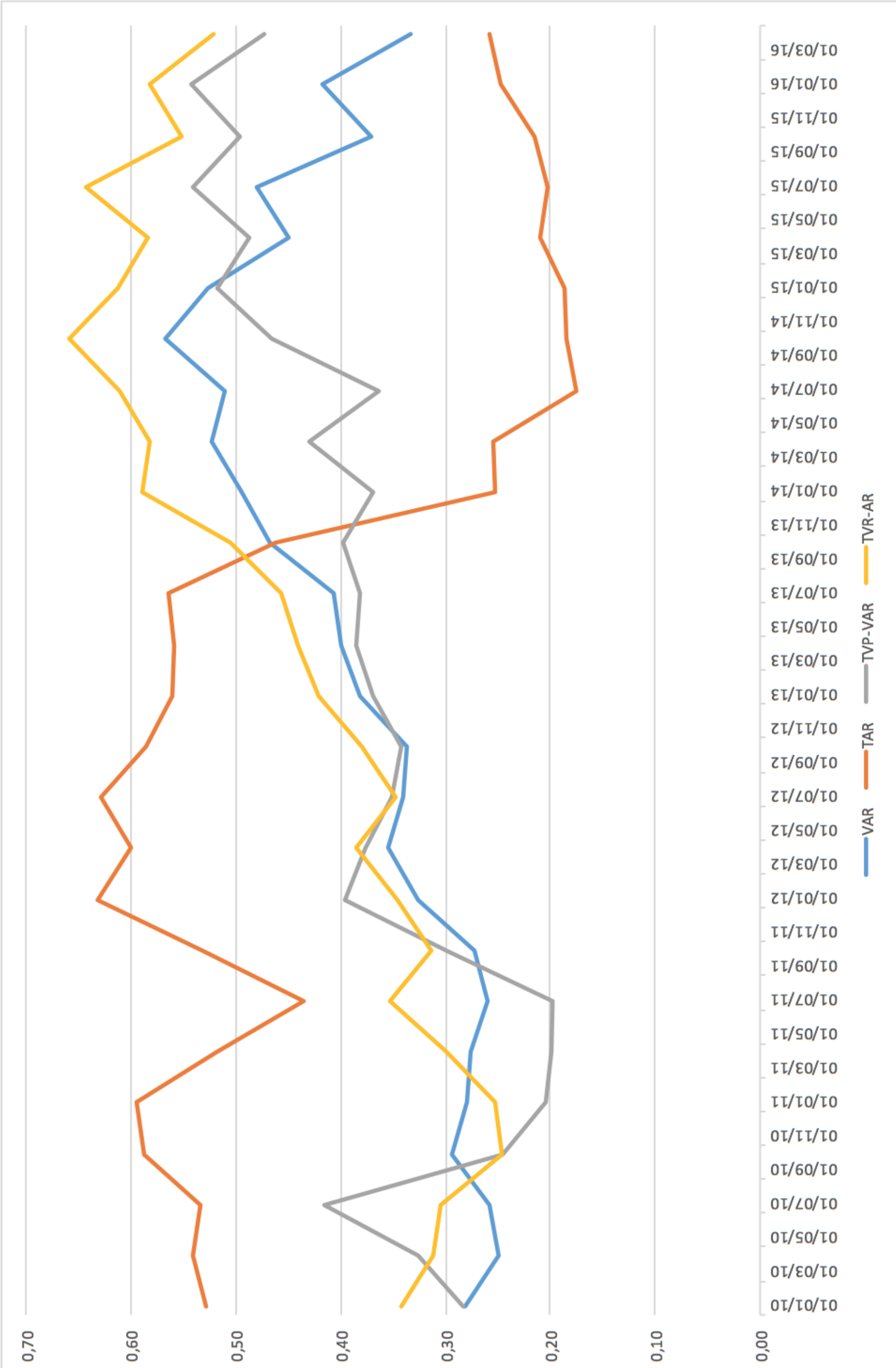

Figure 3.3. RMSE prediction statistics at the eight-quarter horizon for Hungarian inflation

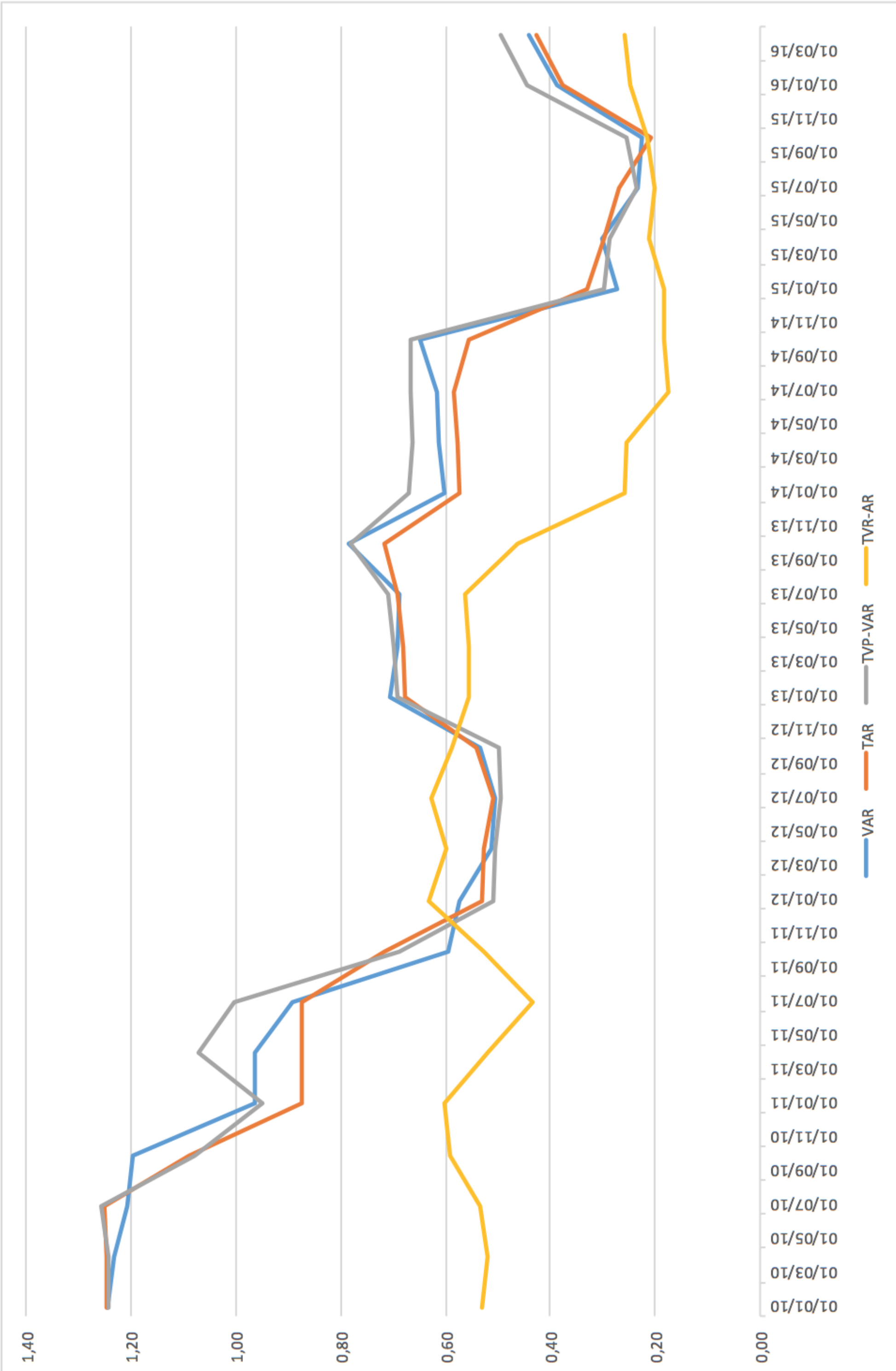

Figure 3.4. RMSE prediction statistics at the eight-quarter horizon for Hungarian output growth

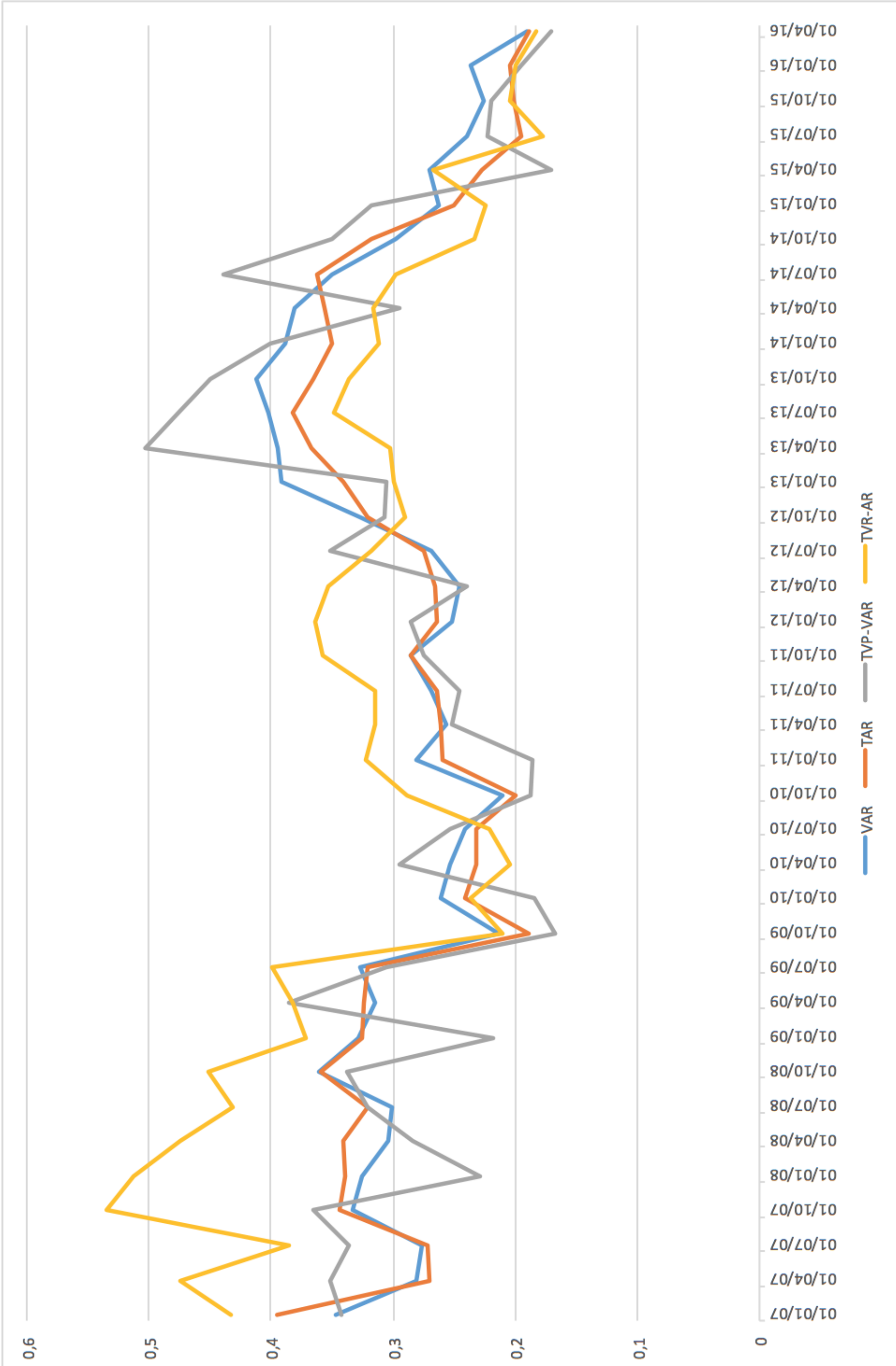

Figure 3.5. RMSE prediction statistics at the eight-quarter horizon for Polish inflation

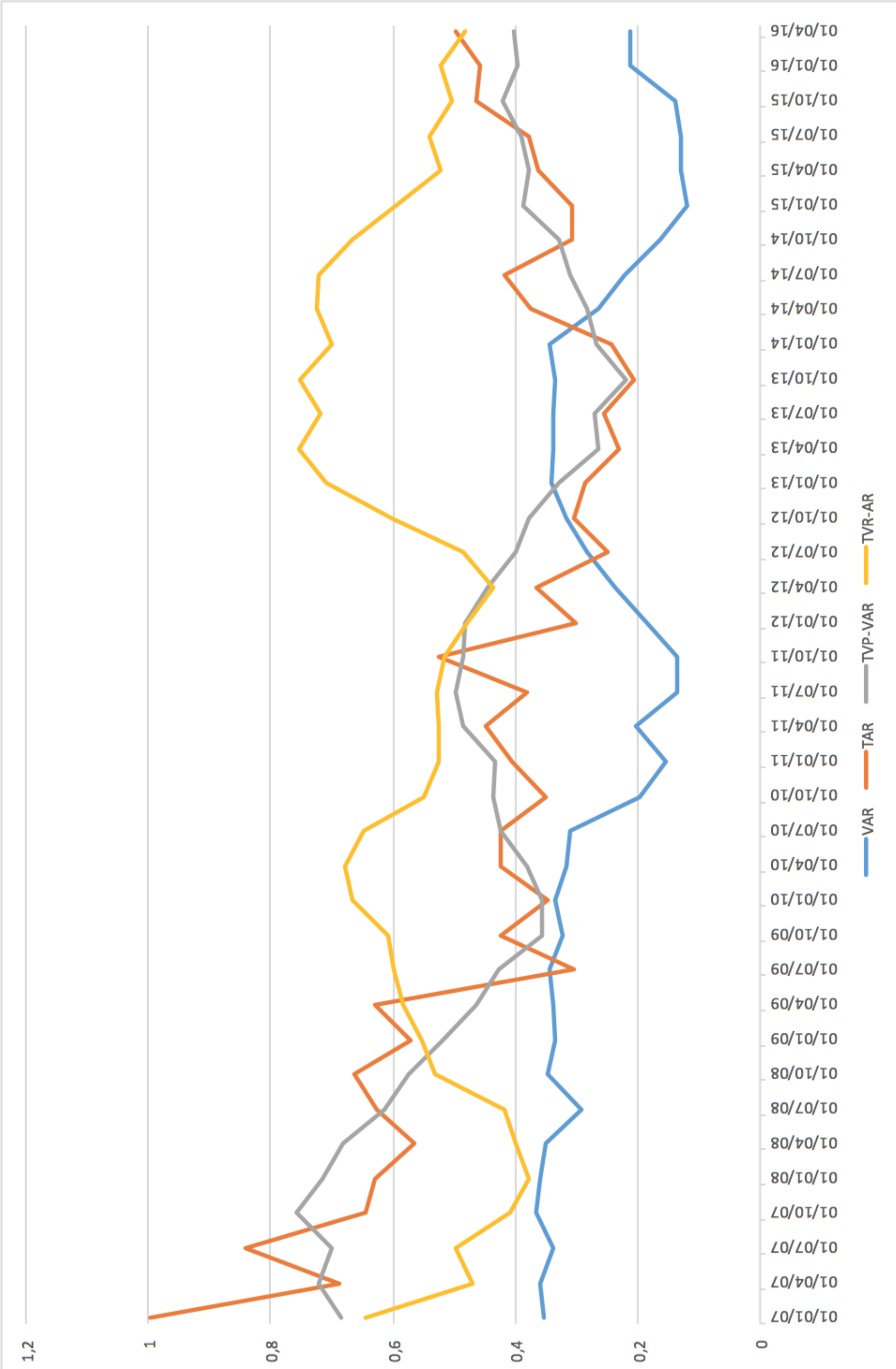

Figure 3.6. RMSE prediction statistics at the eight-quarter horizon for Polish output growth

the Hungarian GDP was declining, the other models perform slightly better.

With regard to Polish data, the models show similar capability to predict the changes in consumer inflation two years ahead, as illustrated by the RMSEs in Figure 3.5. On average, the TVP-VAR performs best, although it is not able to predict the inflation decline that occurred in 2013. Polish GDP growth, whose RMSE are represented in Figure 3.6, is best predicted by an invariant VAR model. This model is outperformed by the TAR and TVP-VAR models only in the period spanning from late 2012 to early 2014, when Polish economy growth slowed down, and the VAR model performance deteriorates. Nevertheless, in forecasting output growth two years ahead in the most recent quarters, it performs quite remarkably.

3.3 Concluding remarks

It seems natural to ask whether identifying structural changes or changes in monetary policy regime could help in improving the accuracy of macroeconomic forecasts. In this chapter, I have compared the forecasting performance of four different econometric models using data of three Eastern European countries. I have estimated the models recursively over an expanding data window. To compare the average forecasting performance of the models, I have chosen two tools, the root mean squared error (RMSE), which evaluates the point forecasts, and the log predictive density scores, which takes the whole densities into account.

To study the possible advantage of allowing for time varying parameters and covariance matrix, I have compared the results following from a TVP-VAR(1) estimation with those resulting from a VAR(2) model with fixed parameters. Fixed parameters VAR models are the most commonly used in the literature, and their advantage in comparison to VAR with time-varying parameters is that a lower number of parameters must be estimated. Thus, given the short data set, this model can be simply estimated with two lags, whereas by time varying models, using more than one lag increases the estimation error significantly, leading to systematically poor results.

My analysis has delivered several interesting results. Perhaps unsurprisingly, the advantages resulting from the inclusion of time varying components are less obvious than in the literature analyzing the time varying models using longer datasets. I have also shown that, as far as Czech and Hungarian data are concerned, the TVP-VAR model performs on average better than the others. However, against Polish data, the VAR(2) model generally performs slightly better. Additionally, when forecasting output growth, the TVP-AR model delivers better forecasting results on average. Last but not least, given a forecast two years ahead, the TAR model is usually less capable to predict troughs than models that allow for time varying parameters.

Conclusion

The thesis has provided detailed empirical applications of two sets of forecasting methods, popular in the academic literature, using macroeconomic time series, separately applied to data of three Eastern European countries (EECs): namely, the Czech republic, Hungary and Poland. The analysis has been a natural extension to my previous studies, particularly those presented in Chapter 2 with regard to the work presented in Junicke (2017), where I applied Bayesian inference to produce an empirical estimation of a dynamic stochastic general equilibrium (DSGE) model for a small open economy.

In the first chapter, I have surveyed of the literature on macroeconomic forecasting. I have introduced the leading forecasting models, separating my analysis of the contributions with theoretical grounds from those with econometric foundations. With regard to the first group, the focus has naturally been on dynamic general equilibrium models of New Keynesian type. The review has shown that there has been a significant improvement of these models in terms of forecasting performance, though the prediction accuracy of alternative models is still generally superior. Nevertheless, DSGE models are considered in the literature as an essential tool for addressing forecasts involving economic policy analysis.

Concerning the second group, I have considered a number of different versions of vector autoregressive models, including VAR with both fixed and time-varying parameters. The study also described structural VARs and the relevant reduced-forms. Naturally, both OLS based and Bayesian VARs have been analyzed. My survey has shown that econometric models are still the best performing in the forecasting arena. Many issues, including the key identification problem, generating criticism towards these models have been dealt with and significantly reduced. Nonetheless, limits appear to still apply when considering complex models, particularly those designed to perform policy analysis. Overall, these models stand as the leader in the macroeconomic forecasting business.

I have also surveyed some studies proposing combinations of the two approaches: the so-called pooled macroeconomic DSGE-VAR and DSGE-DFM models (DFM stands for dynamic factor model). I have then reviewed the contributions that developed the Bayesian inference and forecasting methodologies. I have explored the techniques to measure the models' forecasting accuracy, with reference to both point and density forecast. Finally, I have briefly discussed the scant forecasting applications that use macroeconomic data of the Czech Republic, Hungary and/or

Poland.

In the second chapter, I have developed a simple DSGE model for a small open economy, and studied its forecasting performance. The contribution has been presented as a natural extension of my work in Junicke (2017), where I applied Bayesian estimation techniques to investigate monetary policy in the three countries under consideration using a New Keynesian model. In particular, I have analyzed the forecasting performance of a two-country DSGE model with non zero steady state inflation. First, I have developed a small-scale DSGE model similar to Lubik and Schorfheide (2007), with a micro-founded Phillips curve. I have assumed imperfect pass-through and non-unit intratemporal elasticity of substitution between domestic and foreign goods. I have log-linearized the model around a steady state with non-zero inflation.

I have carried out Bayesian inference, using a Metropolis-Hastings sampling approach, to measure the performance of this model against the Euro area data and. In order to study the model for the SOE, I have used the data of three EECs, namely the Czech Republic, Hungary and Poland. I have shown that the forecasting performance of this model is higher than assuming a benchmark model with zero steady state inflation for all the three countries. This result has been even more remarkable when a recursive forecasting scheme has been used.

In the third chapter, I have studied the performance of several purely econometric linear and non-linear models, namely a linear vector autoregressive model (VAR) with fixed parameters, a threshold vector autoregressive model (TAR), a time varying parameters autoregressive model (TVP-AR) that serves as a benchmark model, and the time-varying parameters vector autoregressive model (TVP VAR). The question that I have posed is whether identifying structural changes or changes in monetary policy regime could help in improving the accuracy of macroeconomic forecasts. I have estimated the models recursively over an expanding data window. To compare the average forecasting performance of the models, I have chosen two tools, the root mean squared error (RMSE), which evaluates the point forecasts, and the log predictive density scores, which takes the whole densities into account.

In order to study the possible advantage of allowing for time varying parameters and covariance matrix, I have compared the results following from a TVP-VAR(1) estimation with those resulting from a VAR(2) model with fixed parameters. Fixed parameters VAR models are the most commonly used in the literature, and their advantage in comparison to VAR with time-varying parameters is that a lower number of parameters must be estimated. Thus, given the short data set, it has been possible to estimate this model with two lags, whereas in the case of time varying models, using more than one lag would have increased the estimation error significantly, leading to systematically poor results.

My analysis has delivered several interesting results. Perhaps unsurprisingly, the advantages resulting from the inclusion of time varying components are less obvious than in the literature analyzing the time varying models using longer datasets. I have also shown that, as far as Czech and Hungarian data are concerned, the TVP-VAR model performs on average better than the others. However, against Polish data, the VAR(2) model generally performs slightly better. Additionally, when forecasting output growth, the TVP-AR model delivers better forecasting results on average. Last but not least, given a forecast two years ahead, the TAR model is usually less capable to predict troughs than models that allow for time varying parameters.

Bibliography

[1] Adolfson, M., Andersson, M.K., Lindé, J., Villani, M. and Vredin, A. (2007). "Modern forecasting models in action: improving macroeconomic analyses at central banks." *International Journal of Central Banking*, 3, 111-144.

[2] Adolfson, M., Laséen, S., Lindé, J., and Villani, M. (2008). "Evaluating an estimated New Keynesian small open economy model." *Journal of Economic Dynamics and Control*, 32, 2690-2721.

[3] Alessandri, P. and Mumtaz, H. (2017). "Financial conditions and density forecasts for U.S. output and inflation." *Review of Economic Dynamics*, 24, 66-78.

[4] Amisano, G. and Geweke, J. (2013). "Prediction using several macroeconomic models." ECB Working Paper Series 1537.

[5] An, S. and Schorfheide, F. (2007). "Bayesian analysis of DSGE models." *Econometric Reviews*, 26, 113-172.

[6] Angeloni, I., Aucremanne, L., Ehrmann, M., Galì, J., Levin, A. and Smets, F. (2004). "Inflation persistence in the euro area: preliminary summary of findings." Manuscript, ECB.

[7] Ascari, G. and Ropele, T. (2007). "Optimal monetary policy under low trend inflation." *Journal of Monetary Economics*, 54, 2568-2583.

[8] Ball, L., Mankiw, N.G. and Reis, R. (2005). "Monetary policy for inattentive economies." *Journal of Monetary Economics*, 52, 703-725.

[9] Bańbura, M., Giannone, D. and Reichlin, L. (2010). "Large Bayesian vector auto regressions." *Journal of Applied Econometrics*, 25, 71-92.

[10] Barnett A., Mumtaz H. and Theodoridis, K. (2014). "Forecasting U.K. GDP growth and inflation under structural change: a comparison of models with time-varying parameters." *International Journal of Forecasting*, 30, 129-143.

[11] Békési, L., Köber, C., Kucsera, H., Várnai, T. and Világi, B. (2016). "The macroeconomic forecasting model of the MNB." . Magyar Nemzeti Bank Working Paper 2016/4.

[12] Benati, L. (2010). "Are policy counterfactuals based on structural VARs reliable?" ECB Working Paper Series 1188.

[13] Benigno, P. and López-Salido, J.D. (2006). "Inflation persistence and optimal monetary policy in the Euro Area." *Journal of Money, Credit & Banking*, 38, 587-615.

[14] Bernanke, B., Gertler, M. and Gilchrist, S. (1999). "The financial accelerator in a quantitative business cycle framework." *Handbook of Macroeconomics*, 1, 1341-1393.

[15] Blake, A. (2012). "Fixed interest rates over finite horizons." Bank of England Working Paper 454.

[16] Blake, M. and Mumtaz, H. (2012) "Applied Bayesian econometrics for central bankers." Bank of England Technical Handbook 4.

[17] Boivin, J. and Giannoni, M.P. (2006). "DSGE models in a data rich environment." NBER w12772.

[18] Brooks, S.P. and Gelman, A. (1998). "General methods for monitoring convergence of iterative simulations." *Journal of Computational and Graphical Statistics*, 7, 434-455.

[19] Brunnermeier, M.K., Eisenbach, T.M. and Sannikov, Y. (2013). "Macroeconomics with Financial Frictions: A Survey." *Advances in Economics and Econometrics: Volume 2, Applied Economics: Tenth World Congress*, 50, 3-94.

[20] Calvo, G.A. (1983). “Staggered prices in a utility-maximizing framework.” *Journal of Monetary Economics*, 12, 383-398.

[21] Canova, F. (2007). G-7 inflation forecasts: random walk, Phillips curve or what else?” *Macroeconomic Dynamics*, 11, 1-30.

[22] Carriero, A., Clark, T. and Marcellino, M. (2011). “Bayesian VARs: specification choices and forecast accuracy.” *Journal of Applied Econometrics*, 30, 46-73.

[23] Chatfield, C. (1993). “Calculating interval forecasts.” *Journal of Business and Economics Statistics*, 11, 121-135.

[24] Chen, C.W.S. (1998). “A Bayesian analysis of generalized threshold autoregressive models.” *Statistics and Probability Letters*, 40, 15-22.

[25] Chen, C.W.S. and Lee, J.C. (1995). “Bayesian inference of threshold autoregressive models.” *Journal of Time Series Analysis*, 16, 483-492.

[26] Chib, S. (1998). “Estimation and comparison of multiple change point models.” *Journal of Econometrics*, 86, 221-241.

[27] Chib, S. and Ramamurthy, S. (2014). “DSGE models with student-t errors.” *Econometric Reviews*, 33, 152-171.

[28] Christiano, L.J., Eichenbaum, M. and Evans, C.L. (2005). "Nominal rigidities and the dynamic effects of a shock to monetary policy." *Journal of Political Economy*, 113, 1-45.

[29] Christiano, L., Motto, R. and Rostagno, M. (2010). "Financial factors in economic fluctuations." ECB Working Papers Series 1192.

[30] Christoffel, K., Coenen, G. and Warne, A. (2011). "Forecasting with DSGE models." *Oxford Handbook of Economic Forecasting*, 89-127.

[31] Clark, T.E. (2011). "Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility." *Journal of Business and Economic Statistics*, 29, 327-341.

[32] Clark, T.E. and McCracken, M.W. (2010). "Averaging forecasts from VARs with uncertain instabilities." *Journal of Applied Econometrics*, 25, 5-29.

[33] Coenen, G., McAdam, P. and Straub, R. (2008). "Tax reform and labor market performance in the Euro Area: a simulation-based analysis using the New Area-Wide model." *Journal of Economic Dynamics and Control*, 32, 2543-2583.

[34] Cogley, T., Morozov S. and Sargent T.J. (2005). "Bayesian fan charts for U.K. inflation: forecasting sources of uncertainty in an evolving monetary system." *Journal of Economic Dynamics and Control*, 29, 1893-1925.

[35] Cogley, T. and Sargent, T.J. (2002). "Evolving post-world war II U.S. inflation dynamics." *NBER Macroeconomics Annual*, 16, 331-388.

[36] Cogley, T. and Sargent, T.J. (2005). "Drifts and volatilities: monetary policies and outcomes in the post WWII U.S." *Review of Economic Dynamics*, 8, 262-302.

[37] Coibion, O., Gorodnichenko, Y. and Wieland, J. (2012). "The optimal inflation rate in New Keynesian models: should central banks raise their inflation targets in light of the zero lower bound?" *Review of Economic Studies*, 79, 1371-1406.

[38] Corsetti, G. and Pesenti, P. (2001). "Welfare and macroeconomic interdependence." *Quarterly Journal of Economics*, 116, 421-445.

[39] Cúrdia, V., Del Negro, M. and Greenwald, D. (2014). "Rare Shocks. Great Recessions." *Journal of Applied Econometrics*, 29, 1031-1052.

[40] D'Agostino, A., Gambetti, L. and Giannone, D. (2013). "Macroeconomic forecasting and structural change." *Journal of Applied Econometrics*, 28, 82-101.

[41] Darvas, Z. (2013). "Monetary transmission in three Central European economies: evidence from time-varying coefficient vector autoregressions." *Empirica*, 40, 363-390.

[42] De Paoli, B. (2009). "Monetary policy and welfare in a small open economy." *Journal of International Economics*, 77, 11-22.

[43] Del Negro, M. and Schorfheide, F. (2004). "Priors from general equilibrium models for VARs." *International Economic Review*, 45, 643-673.

[44] Del Negro, M. and Schorfheide, F. (2009). "Inflation dynamics in a small open economy model under inflation targeting: some evidence from Chile." *Central Banking, Analysis, and Economic Policies Book Series*, 13, 511-562.

[45] Del Negro, M. and Schorfheide, F. (2011). "Bayesian macroeconometrics." *The Oxford handbook of Bayesian econometrics*, 293-389.

[46] Del Negro, M. and Schorfheide, F. (2013). "DSGE model-based forecasting." *Handbook of Economic Forecasting*, 2, 57-140.

[47] Del Negro, M., Schorfheide, F., Smets, F. and Wouters, R. (2007). "On the fit of New Keynesian models." *Journal of Business and Economic Statistics*, 25, 123-162.

[48] Devereux, M.B. and Engel, C. (2003). "Monetary policy in the open economy revisited: price setting and exchange-rate flexibility." *Review of Economic Studies*, 70, 765-783.

[49] Di Giorgio, G. and Nisticò, S. (2007). "Monetary policy and stock prices in an open economy." *Journal of Money, Credit and Banking*, 39, 1947-1985.

[50] Diebold, F.X., Schorfheide, F. and Shin, M. (2016). "Real-time forecast evaluation of DSGE models with stochastic volatility." NBER w22615.

[51] Doan, T., Litterman, R.B. and Sims, C.A. (1984). "Forecasting and conditional projection using realistic prior distributions." *Econometric Reviews*, 3, 1-144.

[52] Dupuis, D. (2004). "The New Keynesian hybrid Phillips curve: An assessment of competing specifications for the United States." Bank of Canada Working Papers 04-31.

[53] Edge, R. and Gürkaynak, R. (2010). "How useful are estimated DSGE model forecasts for central bankers?" *Brookings Papers on Economic Activity*, 41, 209-259.

[54] Edge, R.M., Kiley, M.T. and Laforte, J.P. (2010). "A comparison of forecast performance between Federal Reserve staff forecasts, simple reduced-form models, and a DSGE model." *Journal of Applied Econometrics*, 25, 720-754.

[55] Eichenbaum, M. and Fisher, J.D. (2003). "Evaluating the Calvo model of sticky prices." NBER w10617.

[56] Fernández-Villaverde, J. and Rubio-Ramírez, J.F. (2013). "Macroeconomics and volatility: data, models, and estimation." *Advances in Economics and Econometrics: Tenth World Congress*, 3, 137-183.

[57] Franta, M., Horvath, R. and Rusnak, M. (2014). "Evaluating changes in the monetary transmission mechanism in the Czech Republic." *Empirical Economics*, 46, 827-842.

[58] Fuhrer, J.C. (2000)."Habit formation in consumption and its implications for monetary-policy models." *American Economic Review*, 90, 367-390.

[59] Galì, J. and Gertler, M. (1999). Inflation dynamics: a structural econometric analysis." *Journal of Monetary Economics*, 44, 195-222.

[60] Galì, J., Gertler, M. and Lopez-Salido, J.D. (2001). "European inflation dynamics." *European Economic Review*, 45, 1237-1270.

[61] Galì, J. and Monacelli, T. (2005). "Monetary policy and exchange rate volatility in a small open economy." *Review of Economic Studies*, 72, 707-734.

[62] George, E.I., Sun, D. and Ni, S. (2008). "Bayesian stochastic search for VAR model restrictions." *Journal of Econometrics*, 142, 553-580.

[63] Gerard, H. and Nimark, K. (2008). "Combining multivariate density forecasts using predictive criteria." Reserve Bank of Australia Working Paper rdp2008-02.

[64] Geweke, J. (1977). "The dynamic factor analysis of economic time series models." *Latent variables in socio-economic models*, 19, 365-383.

[65] Ghent, A.C. (2009). "Comparing DSGE-VAR forecasting models: how big are the differences?" *Journal of Economic Dynamics and Control*, 33, 864-882.

[66] Giannone, D., Lenza, M. and Primiceri, G.E. (2015). "Prior selection for vector autoregressions." *Review of Economics and Statistics*, 97, 436-451.

[67] Giorgio, G.D. and Nistico, S. (2007). "Monetary policy and stock prices in an open economy." *Journal of Money, Credit and Banking*, 39, 1947-1985.

[68] Greene, W.H. (2003). *Econometric Analysis*. Prentice Hall.

[69] Hamilton, J.D. (1994). *Time series analysis*, 2. Princeton.

[70] Herbst, E. and Schorfheide, F. (2012). "Evaluating DSGE model forecasts of comovements." *Journal of Econometrics*, 171, 152-166.

[71] Highfield, R.A. (1987). "Forecasting with Bayesian state space models." PhD thesis, University of Chicago.

[72] Hirano, K. and Wright, J.H. (2017). "Forecasting with model uncertainty: representations and risk reduction." *Econometrica*, 85, 617-643.

[73] Ireland, P.N. (2007). "Changes in the Federal Reserve's inflation target: causes and consequences." *Journal of Money, Credit and Banking*, 39, 1851-1882.

[74] Jochmann, M., Koop, G. and Strachan, R. (2010). "Bayesian forecasting using stochastic search variable selection in a VAR subject to breaks." *International Journal of Forecasting*, 26, 326-347.

[75] Junicke, M. (2017). "Trend inflation and monetary policy in Eastern Europe." *Macroeconomic Dynamics*, forthcoming.

[76] Junicke, M. and Merella, V. (2017). "The impact of foreign monetary policy on Eastern European countries." Manuscript, University of Cagliari.

[77] Justiniano, A. and Preston, B. (2010). "Monetary policy and uncertainty in an empirical small open-economy model." *Journal of Applied Econometrics*, 25, 93-128.

[78] Justiniano, A. and Primiceri, G.E. (2008). "The time-varying volatility of macroeconomic fluctuations." *American Economic Review*, 98, 604-641.

[79] Kadiyala, K.R. and Karlsson, S. (1997). "Numerical methods for estimation and inference in Bayesian VAR models." *Journal of Applied Econometrics*, 12, 99-132.

[80] Karlsson, S. (2013). "Forecasting with Bayesian vector autoregressions." *Handbook of economic forecasting*, 2, 791-897.

[81] Kim, C.J. and Nelson, C.R. (1999). "Has the U.S. economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle." *Review of Economics and Statistics*, 81, 608-616.

[82] King, R.G., Plosser, C.I. and Rebelo, S.T. (1988). "Production, growth and business cycles: I. the basic neoclassical model." *Journal of Monetary Economics*, 21, 195-232.

[83] Kolasa, M., Rubaszek, M. and Skrzypczyński, P. (2012). "Putting the New Keynesian DSGE model to the real-time forecasting test." *Journal of Money, Credit and Banking*, 44, 1301-1324.

[84] Koop, G. (2003). *Bayesian Econometrics.* Wiley.

[85] Koop, G. and Korobilis, D. (2009). "Bayesian multivariate time series methods for empirical macroeconomics." *Foundations and Trends in Econometrics*, 3, 267-358.

[86] Koop, G. and Potter, S. (2007). "Estimation and forecasting in models with multiple breaks." *Review of Economic Studies*, 74, 763-789.

[87] Korobilis, D. (2008). "Forecasting in vector autoregressions with many predictors." *Bayesian Econometrics*, 403-431.

[88] Korobilis, D. (2013). "VAR forecasting using Bayesian variable selection." *Journal of Applied Econometrics*, 28, 204-230.

[89] Kozicki, S. and Tinsley, P.A. (2002). "Dynamic specifications in optimizing trend-deviation macro models." *Journal of Economic Dynamics and Control*, 26, 1585-1611.

[90] Kryshko, M. (2011). "Data-rich DSGE and dynamic factor models." IMF Working Paper 11/216.

[91] Kydland, F.E. and Prescott, E.C. (1982). "Time to build and aggregate fluctuations." *Econometrica*, 50, 1345-1370.

[92] Laseen, S. and Svensson, L.E. (2011). "Anticipated alternative policy-rate paths in policy simulations." *International Journal of Central Banking*, 7, 1-35.

[93] Leeper, E.M. and Zha, T. (2003). "Modest policy interventions." *Journal of Monetary Economics*, 50,1673-1700.

[94] Levine, P., Pearlman, J., Perendia, G. and Yang, B. (2012). "Endogenous persistence in an estimated DSGE model under imperfect information, *Economic Journal*, 122, 1287-1312.

[95] Litterman, R.B. (1979). "Techniques of forecasting using vector autoregressions." Federal Reserve Bank of Minneapolis Working Paper 115.

[96] Litterman, R.B. (1986). "Forecasting with Bayesian vector autoregressions—five years of experience." *Journal of Business and Economic Statistics*, 4, 25-38.

[97] Liu, P. (2006). "A small New Keynesian model of the New Zealand economy." Reserve Bank of New Zealand Discussion Paper Series DP2006/03.

[98] Lopes, H.F. and Salazar, E. (2006). "Bayesian model uncertainty in smooth transition autoregressions." *Journal of Time Series Analysis*, 27, 99-117.

[99] Lubik, T. and Schorfheide, F. (2005). "A Bayesian look at new open economy macroeconomics." *NBER Macroeconomics Annual*, 20, 313-366.

[100] Lubik, T. and Schorfheide, F. (2007). "Do central banks respond to exchange rate movements? A structural investigation." *Journal of Monetary Economics*, 54, 1069-1087.

[101] Lucas, R.J. (1976). "Econometric policy evaluation: A critique." *Carnegie-Rochester Conference Series on Public Policy*, 1, 19-46.

[102] Maćkowiak, B. (2006). "How much of the macroeconomic variation in Eastern Europe is attributable to external shocks?" *Comparative Economic Studies*, 48, 523-544.

[103] Mandalinci, Z. (2015). "Forecasting inflation in emerging markets: an evaluation of alternative models." CReMFi Discussion Paper 3.

[104] Marcellino, M. and Rychalovska, Y. (2014). "Forecasting with a DSGE model of a small open economy within the monetary union." *Journal of Forecasting*, 33, 315-338.

[105] McNees, S.K. (1986). "Forecasting accuracy of alternative techniques: a comparison of U.S. macroeconomic forecasts. " *Journal of Business and Economic Statistics*, 4, 5-15.

[106] McNees, S.K. (1990). "The role of judgment in macroeconomic forecasting accuracy." *International Journal of Forecasting*, 6, 287-299.

[107] Milani, F. and Treadwell, J. (2012). "The effects of monetary policy 'news' and 'surprises'." *Journal of Money, Credit and Banking*, 44, 1667-1692.

[108] Monacelli, T. (2003). "Monetary policy in a low pass-through environment." ECB Working Paper Series 227.

[109] Mumtaz, H. and Sunder-Plassmann, L. (2013). "Time-varying dynamics of the real exchange rate: an empirical analysis." *Journal of Applied Econometrics*, 28, 498-525.

[110] Obstfeld, M. and Rogoff, K. (2002). "Global implications of self-oriented national monetary rules." *Quarterly Journal of Economics*, 117, 503-535.

[111] Österholm, P. (2008). A structural Bayesian VAR for model-based fan charts." *Applied Economics*, 40, 1557-1569.

[112] Pesaran, M.H., Petenuzzo, D. and Timmermann, A. (2006). "Forecasting time series subject to multiple structural breaks." *Review of Economic Studies*, 73, 1057-1084.

[113] Phillips, A.W. (1958). "The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957." *Economica*, 25, 283-299.

[114] Pichler, P. (2008). "Forecasting with DSGE models: the role of nonlinearities." *B.E. Journal of Macroeconomics*, 8, 1-35.

[115] Primiceri, G.E. (2005). "Time varying structural vector autoregressions and monetary policy." *Review of Economic Studies*, 72, 821-852.

[116] Rabanal, P. and Rubio-Ramírez, J.F. (2015). "Can international macroeconomic models explain low-frequency movements of real exchange rates?" *Journal of International Economics*, 96, 199-211.

[117] Rabanal, P. and Tuesta, V. (2010). "Euro-dollar real exchange rate dynamics in an estimated two-country model: an assessment." *Journal of Economic Dynamics and Control*, 34, 780-797.

[118] Roberts, J.M. (2001). "How well does the New Keynesian sticky-price model fit the data?" *Contributions in Macroeconomics*, 5, 1-37.

[119] Rotemberg, J.J. (1982). "Sticky prices in the United States." *Journal of Political Economy*, 90, 1187-1211.

[120] Sargent, T.J. (1989). "Two models of measurements and the investment accelerator." *Journal of Political Economy*, 97, 251-287.

[121] Sargent, T.J., and Sims, C.A. (1977). "Business cycle modeling without pretending to have too much *a priori* economic theory." *New Methods in Business Cycle Research*, 1, 145-168.

[122] Sbordone, A.M. (2002). "Prices and unit labor costs: a new test of price stickiness." *Journal of Monetary Economics*, 49, 265-292.

[123] Schorfheide, F. (2008). "DSGE model-based estimation of the New Keynesian Phillips curve." *Economic Quarterly*, 94, 397-433.

[124] Schorfheide, F., Sill, K. and Kryshko, M. (2010). "DSGE model-based forecasting of non-modelled variables." *International Journal of Forecasting*, 26, 348-373.

[125] Sims, C.A. (1980). "Macroeconomics and reality." *Econometrica*, 48, 1-48.

[126] Sims, C.A. (1993). "A nine-variable probabilistic macroeconomic forecasting model." *Business Cycles, Indicators and Forecasting*, 28, 179-204.

[127] Sims, C.A. and Zha, T. (1998). "Bayesian methods for dynamic multivariate models." *International Economic Review*, 39, 949-968.

[128] Smets, F. and Wouters, R. (2003). "An estimated dynamic stochastic general equilibrium model of the euro area." *Journal of the European Economic Association*, 1, 1123-1175.

[129] Smets, F. and Wouters, R. (2005). "Bayesian new neoclassical synthesis (NNS) models: modern tools for central banks." *Journal of the European Economic Association*, 3, 422-433.

[130] Smets, F. and Wouters, R. (2007). "Shocks and frictions in U.S. business cycles: a Bayesian DSGE approach." *American Economic Review*, 97, 586-606.

[131] Steinbach, M., Mathuloe, P. and Smit, B. (2009). "An open economy New Keynesian DSGE model of the South African economy." *South African Journal of Economics*, 77, 207-227.

[132] Stock, J.H. and Watson, M.W. (1996). "Evidence on structural instability in macroeconomic time series relations." *Journal of Business and Economic Statistics*, 14, 11-30.

[133] Stock, J.H. and Watson, M.W. (2001). "Vector autoregressions." *Journal of Economic Perspectives*, 15, 101-115.

[134] Stock, J.H. and Watson, M.W. (2011). "Dynamic factor models." *Oxford Handbook on Economic Forecasting*, 35-60.

[135] Sutherland, A. (2002). "A simple second-order solution method for dynamic general equilibrium models." Department of Economics Discussion Paper Series 200211, University of St. Andrews.

[136] Sutherland, A. (2006). "The expenditure switching effect, welfare and monetary policy in a small open economy." *Journal of Economic Dynamics and Control*, 30, 1159-1182.

[137] Taylor, J.B. (1983). "Union wage settlements during a disinflation." *American Economic Review*, 73, 981-993.

[138] Villani, M. (2001). "Bayesian prediction with cointegrated vector autoregressions." *International Journal of Forecasting*, 17, 585-605.

[139] Villani, M. (2009). "Steady state priors for vector autoregressions." *Journal of Applied Econometrics*, 24, 630-650.

[140] Waggoner, D.F. and Zha, T. (2012). "Confronting model misspecification in macroeconomics." *Journal of Econometrics*, 171, 167-184.

[141] Warne, A., Coenen, G. and Christoffel, K. (2014). "Predictive likelihood comparisons with DSGE and DSGE-VAR models." CFS Working Paper 478.

[142] Wieland, V. and Wolters, M.H. (2011). "The diversity of forecasts from macroeconomic models of the U.S. economy." *Economic Theory*, 47, 247-292.

[143] Wieland, V. and Wolters, M.H. (2013). "Forecasting and policy making." *Handbook of economic forecasting*, 2, 239-325.

[144] Wolters, M.H. (2011). "Forecasting under model uncertainty." German Economic Association Working Paper 48723.

[145] Zarnowitz, V. and Braun, P. (1993). "Twenty-two years of the NBER-ASA quarterly economic outlook surveys: aspects and comparisons of forecasting performance." *Business Cycles,Indicators, and Forecasting*, 28, 11-94.

[146] Zellner, A. (1971). *An introduction to Bayesian inference in econometrics.* Wiley.