Title	On the predictability of time-varying VAR and DSGE models
Authors(s)	Bekiros, Stelios D., Paccagnini, Alessia
Publication date	2013-08
Publication information	Bekiros, Stelios D., and Alessia Paccagnini. “On the Predictability of Time-Varying VAR and DSGE Models” 45, no. 1 (August, 2013).
Publisher	Springer
Item record/more information	http://hdl.handle.net/10197/7329
Publisher's statement	The final publication is available at www.springerlink.com.
Publisher's version (DOI)	10.1007/s00181-012-0623-z

Noname manuscript No.
(will be inserted by the editor)

On the predictability of time-varying VAR and DSGE models

Stelios Bekiros · Alessia Paccagnini

Received: date / Accepted: date

Abstract Over the last few years, there has been a growing interest in DSGE modelling for predicting macroeconomic fluctuations and conducting quantitative policy analysis. Hybrid DSGE models have become popular for dealing with some of the DSGE misspecifications as they are able to solve the trade-off between theoretical coherence and empirical fit. However, these models are still linear and they do not consider time-variation for parameters. The time-varying properties in VAR or DSGE models capture the inherent non-linearities and the adaptive underlying structure of the economy in a robust manner. In this paper, we present a state space time-varying parameter VAR model. Moreover, we focus on the DSGE-VAR that combines a micro-founded DSGE model with the flexibility of a VAR framework. All the aforementioned models as well simple DSGEs and Bayesian VARs are used in a comparative investigation of their out-of-sample predictive performance regarding the US economy. The results indicate that while in general the classical VAR and BVARs provide with good forecasting results, in many cases the TVP-VAR and the DSGE-VAR outperform the other models.

Keywords Hybrid DSGE · Time-varying VAR · Kalman filter · Bayesian VAR · Forecasting

JEL Classification C11 · C15 · C32

S. Bekiros
European University Institute (EUI), Department of Economics, Via della Piazzuola 43 I-50133, Florence, Italy
Tel.: +39-055-4685-916
Fax: +39-055-4685-902
E-mail: stelios.bekiros@eui.eu

A. Paccagnini
Università degli Studi di Milano-Bicocca, Department of Economics, Piazza Ateneo Nuovo 1, 20126 Milano, Italy and European University Institute (EUI), Max Weber Programme Via delle Fontanelle 10, I-50014 Florence, Italy
E-mail: alessia.paccagnini@unimib.it

1 Introduction

The standard econometric approach in evaluating the empirical performance of economic models involves the use of statistical methods (Watson 1993). However, there is a growing controversy in the literature about the appropriate empirical methodology in macroeconomic modelling. Since the early 1980s, two streams have emerged in the literature. The first is based on vector autoregressive (VAR) modelling introduced by Sims (1980). VAR models can be applied directly on the data to perform statistical hypothesis (Sims 1982, 1998; Stock and Watson 2001). The VAR is a powerful tool for empirical validation of macroeconomic models, since it is essentially an easy statistical model to estimate and once identification restrictions are imposed, it can be used to evaluate the impact of economic shocks on key variables. Litterman (1986) also used VAR models for forecasting. Nevertheless, even though the VAR model is proven to be a reliable tool in terms of data description and forecasting the classical VAR modelling fails to take into account the inherent nonlinearities of the economy.

The second approach, was initiated by Kydland and Prescott (1982) and Long and Plosser (1983), and became increasingly popular for evaluating dynamic macroeconomic models. Dynamic stochastic general equilibrium (DSGE) models describe the general equilibrium of a model economy in which agents (e.g., consumers, firms etc.) maximize their objectives subject to budget and resource constraints (Del Negro and Schorfheide 2003). The structural parameters of the DSGE model in principle do not vary according to the policy regime. Over the last few years, there has been a growing interest in academia and in central banks in using DSGE models to explain macroeconomic fluctuations and conduct quantitative policy analysis. DSGE models have the advantage of combining the micro-foundations of both households and firm optimization problems with price and wage rigidities. Model validation using DSGE models allows the econometrician to establish a link between structural features of the economy and reduced form parameters, something that was not always possible with the usual large-scale macroeconomic models. Improvements in computational power and the development of new econometric methods are crucial to the popularity of the use of DSGE models. The combination of rich structural models, novel solution algorithms and powerful simulation techniques has allowed researchers to develop the so-called "New Macro-econometrics" (Fernandez-Villaverde 2009). However, the calibrated DSGE models are typically too stylized to be taken directly to the data and often yield weak results (Stock and Watson 2001; Ireland 2004). Despite their recent popularity, DSGE face many important challenges. For instance, Schorfheide (2010) reports five main challenges, namely the fragility of parameter estimates, lack of distinction between exogenous shocks that capture aggregate uncertainty or possible misspecification, the presence of trends, the statistical fit and the weak reliability of policy predictions. The five challenges discussed by Schorfheide (2010) are not the only problems in using DSGE models. Sometimes DSGEs exhibit nonlinearities, even if the common practice is to solve and estimate a linearized version with Gaussian shocks. A number of papers report the lack of possible stochastic volatility or parameter drifts using ad-hoc DSGE models applied to the Great Moderation period such as Kim and Nelson (1999), McConnell and Pérez-Quirós (2000), Clarida *et al.* (2000), Lubik and Schorfheide (2004), Canova and Gambetti (2004), Primiceri (2005), Cogley and Sargent (2005), Sims and Zha (2006), Justiniano and Primiceri (2008) and Benati and Surico (2009), among the most cited.

DSGE models were not considered as forecasting tools until very recent years, when Smets and Wouters (2003, 2004) presented an interesting study of the forecasting performance of DSGE models compared to alternative non-structural models. Moreover, very few papers discuss DSGE model validation, despite its recent use for forecasting (Edge and Gürkaynak 2011). In the very recent macro-econometric literature, hybrid or mixture models have become popular for dealing with some of the DSGE model misspecifications. These models are able to solve the trade-off between theoretical coherence and empirical fit. Essentially, two approaches exist in building empirical models that combine the restrictions of a DSGE model with a pure statistical model (Schorfheide 2010). These are additive hybrid models and hierarchical hybrid models. The hybrid models provide a complete analysis of the law of motion of the data, capturing the dynamic properties of the DSGE model. Different attempts of hybrid models have been introduced for solving, estimating and forecasting with the DSGE model. Sargent (1989) and Altug (1989) proposed augmenting a DSGE model with measurement error terms following a first order autoregressive, known as the DSGE-AR approach. Ireland (2004) proposed a method that is similar to the DSGE-AR, but imposing no restriction on the measurement errors, assuming that residuals follow a first-order vector autoregression (DSGE-AR à l'Ireland). A different approach called DSGE-VAR was proposed by Del Negro and Schorfheide (2004) and was based on the works DeJong *et al.* (1996) and Ingram and Whiteman (1994). The main idea behind the DSGE-VAR is the use of the VAR representation as an econometric tool for empirical validation, combining prior information derived from the DSGE model in estimation. However, it has several problems. One of the main problems in finding a statistical representation for the data by using a VAR is "overfitting" due to the inclusion of too many lags and too many variables, some of which may be insignificant. The problem of "overfitting" results in multicollinearity and the loss of degrees of freedom, leading to inefficient estimates and large out-of-sample forecasting errors. It is possible to overcome this problem by using what have become well-known as "Minnesota" priors (Doan *et al.* 1984). The use of Minnesota priors has been proposed to shrink the parameters space and thus overcome the curse of dimensionality. Overall, there are several examples of additive hybrid models: the DSGE-AR (Sargent 1989; Altug 1989), the DSGE-AR à l'Ireland (2004), the DSGE-DFM by Boivin and Giannoni (2006) and Kryshko (2010) etc. Also, there are some examples of hierarchical hybrid models, such as the well-known DSGE-VAR of Del Negro and Schorfheide (2004) and the Augmented (B)VAR by Fernández-de-Córdoba and Torres (2010). However, these models are still linear and they do not consider time-variation for parameters.

In general, the classical VAR or DSGE modelling fails to take into account the inherent nonlinearities of the economy. In these cases, time-varying parameters or adaptive modelling seem to be attractive alternatives. The time-varying properties are very useful, because they relax stationarity assumptions, and they also provide a simple interpretation of functional coefficients. Time varying autoregressive (TVP-VAR) models have been developed since the early 1980's. Prado and West (2001) offers an excellent review. Primiceri (2005) used them extensively in analyzing macroeconomic policy issues. All parameters in the TVP-VAR specification are assumed to follow the first-order random walk process, thus allowing both temporary and permanent shift in the parameters. Time varying VAR models led to new methods of time series decomposition and analysis as presented with applications in Primiceri (2005). Dahlhaus (1997, 2000) developed asymptotic estimators and results. Time-varying VARs put quite a challenge on an econometrician because of the amount of parameters to estimate. While it is possible to analytically produce the likelihood for the estimation problem, it is frequently difficult to maximize it over such a high dimension. Bayesian estimation with informative or diffuse priors can be considered a way to tackle the problem.

In this paper, we propose a novel time-varying multivariate state-space estimation method for TVP-VAR processes. The state space model has been well studied by Harvey (1990) and Durbin and Koopman (2002). For the TVP-VAR model, the parameters are estimated using a multivariate specification of the standard Kalman filter (Harvey 1990). The likelihood estimation of the TVP-VAR is performed with a suitable multivariate extension of the Kim and Nelson (1999) method. Moreover, we focus on two DSGE models, i.e., the simple DSGE and the DSGE-VAR model. This hybrid DSGE model combines a micro-founded DSGE model with the flexibility of a VAR framework. All the aforementioned models as well standard VARs and Bayesian VARs, are used in a comparative investigation of their out-of-sample predicting performance regarding the US economy. The motivation comes from a group of recent papers that compares the forecasting performance of DSGE against VAR models. This includes Smets and Wouters (2004), Ireland (2004), Del Negro and Schorfheide (2004), Del Negro *et al.* (2007), Adolfson *et al.* (2008), Christoffel *et al.* (2008), Rubaszek and Skrzypczynski (2008), Ghent (2009), Kolasa *et al.* (2009), Consolo *et al.* (2009), Wang (2009), Fernandez-de-Cordoba and Torres (2010), among others. A general result is that the use of the simple or hybrid DSGE improves the forecasting performance compared to VAR and BVAR models. From these studies, we picked out the most commonly used models in order to perform a comparative evaluation. We selected the current models exactly because they are indicative and at the same time representative of the VAR and DGSE classes. Moreover, this choice was motivated by the fact that we are concerned with the three key macroeconomic variables that appear as observables in the simple DGSE model of Del Negro *et al.* (2004). In this study, the GDP, CPI and interest rate forecasts for the US economy derived from the TVP-VAR, DSGE and DSGE-VAR hybrid model are compared with each other and against the forecasts generated from the classical and Bayesian variants of the VAR for a total period of 1980:1 to 2009:4 including the out-of-sample testing period 2001:1-2009:4.

The remainder of this paper is organized as follows. Section 2 describes the DSGE models and the hybrid DSGE-VAR along with the standard classical and Bayesian VARs. Section 3 presents the time-varying multivariate state-space TVP-VAR model. In section 4 the data are described and the empirical results of the comparative predictive investigation are illustrated and analyzed. Finally, section 5 concludes.

2 DSGE Modeling

DSGE models have been considered as forecasting tools only since the seminal work of Smets and Wouters (2003, 2004). Calibrated DSGE models often yield fragile results, when traditional econometric methods are used for estimation (Smets and Wouters 2003; Ireland 2004). Following this idea of combining the DSGE model information and the VAR representation, among other models that have been proposed in the literature, in this study we use the DSGE-VAR hybrid model.

2.1 Simple DSGE model

The simple DSGE model with forward-looking features is usually referred to as a benchmark in the literature. For instance, Del Negro and Schorfheide (2004) used this model to introduce the DSGE-VAR, and investigate its predictive ability. Wang (2009) proposes the same model in another forecasting exercise without using the VAR representation of the DSGE model. In a DSGE setup the economy is made up of four components. First component is the representative household with habit persistent preferences. This household maximizes an additively separable utility function which is separable into consumption, real money balances and hours worked over an infinite lifetime. The household gains utility from consumption and earns interest from holding government bonds and real profits from the firms. It also pays lump-sum taxes to the government. The second component is a perfectly competitive, representative final goods producer which is assumed to use a continuum of intermediate goods as inputs, and the prices for these inputs are given. The producers of these intermediate goods are monopolistic firms with the same production technology. Nominal rigidities are introduced in terms of price adjustment costs for the firms. Each firm maximizes its profits over an infinite lifetime by choosing its labour input and its price. The third component is the government which spends in each period a fraction of the total output that fluctuates exogenously. The government issues bonds and levies lump-sum taxes, which are the main part of its budget constraint. The last component is the monetary authority, which follows a Taylor rule regarding the inflation target and the output gap. There are three economic shocks: an exogenous monetary policy shock (in the monetary policy rule), and two autoregressive processes, AR(1), which model government spending and technology shocks.

To solve the model, optimality conditions are derived for the maximization problems. After linearization around the steady-state, the economy is described by the following system of equations

$$\tilde{x}_t = E_t[\tilde{x}_{t+1}] - \frac{1}{\tau}(\tilde{R}_t - E_t[\tilde{\pi}_{t+1}]) + (1-\rho_g)\tilde{g}_t + \rho_Z \frac{1}{\tau}\tilde{z}_t \tag{1}$$

$$\tilde{\pi}_t = \beta E_t[\tilde{\pi}_{t+1}] + \kappa[\tilde{x}_t - \tilde{g}_t] \tag{2}$$

$$\tilde{R}_t = \rho_R \tilde{R}_{t-1} + (1-\rho_R)(\psi_1 \tilde{\pi}_t + \psi_2 \tilde{x}_t) + \epsilon_{R,t} \tag{3}$$

$$\tilde{g}_t = \rho_g \tilde{g}_{t-1} + \epsilon_{g,t} \tag{4}$$

$$\tilde{z}_t = \rho_z \tilde{z}_{t-1} + \epsilon_{z,t}, \tag{5}$$

where x is the detrended output (divided by the non-stationary technology process), π is the gross inflation rate, and R is the gross nominal interest rate. The tilde denotes percentage deviations from a steady state or, in the case of output, from a trend path (King 2000; Woodford 2003). The model can be solved by applying the algorithm proposed by Sims (2002). Define the vector of variables $\tilde{\mathbf{Z}}_t = \left(\tilde{x}_t, \tilde{\pi}_t, \tilde{R}_t, \tilde{g}_t, \tilde{z}_t, E_t\tilde{x}_{t+1}, E_t\tilde{\pi}_{t+1}\right)$ and the vector of shocks as $\epsilon_t = (\epsilon_{R,t}, \epsilon_{g,t}, \epsilon_{z,t})$. Therefore the previous set of equations, (1) - (5), can be recasted into a set of matrices ($\mathbf{\Gamma}_0, \mathbf{\Gamma}_1, \mathbf{C}, \mathbf{\Psi}, \mathbf{\Pi}$) accordingly to the definition of the vectors $\tilde{\mathbf{Z}}_t$ and ϵ_t

$$\mathbf{\Gamma}_0 \tilde{\mathbf{Z}}_t = \mathbf{C} + \mathbf{\Gamma}_1 \tilde{\mathbf{Z}}_{t-1} + \mathbf{\Psi}\epsilon_t + \mathbf{\Pi}\eta_t \tag{6}$$

where C is a vector of constants, ϵ_t is an exogenously evolving random disturbance and η_t is a vector of expectations errors, $\left(E_t\left(\eta_{t+1}\right) = \mathbf{0}\right)$, not given exogenously but to be treated as part of the model solution. In order to provide the mapping between the observable data and those computed as deviations from the steady state of the model we set the following measurement equations as in Del Negro and Schorfheide (2004)

$$\begin{aligned} \Delta \ln x_t &= \ln \gamma + \Delta \tilde{x}_t + \tilde{z}_t \\ \Delta \ln P_t &= \ln \pi^* + \tilde{\pi}_t \\ \ln R_t^a &= 4\left[(\ln r^* + \ln \pi^*) + \tilde{R}_t\right] \end{aligned} \tag{7}$$

which can be also casted into matrices as

$$\mathbf{Y}_t = \mathbf{\Lambda}_0(\theta) + \mathbf{\Lambda}_1(\theta)\tilde{\mathbf{Z}}_t + v_t \tag{8}$$

where $\mathbf{Y}_t = (\Delta \ln x_t, \Delta \ln P_t, \ln R_t)'$, $v_t = 0$ and $\mathbf{\Lambda}_0$ and $\mathbf{\Lambda}_1$ are defined accordingly. For completeness, we write the matrices **T, R,** $\mathbf{\Lambda}_0$ and $\mathbf{\Lambda}_1$ as a function of the structural parameters in the model, $\theta = \begin{pmatrix} \ln\gamma, \ln\pi^*, \ln r^*, \kappa, \tau, \psi_1, \psi_2, \\ \rho_R, \rho_g, \rho_Z, \sigma_R, \sigma_g, \sigma_Z \end{pmatrix}'$. Such a formulation derives from the rational expectations solution. The evolution of the variables of interest, $\mathbf{Y}_t$, is therefore determined by (6) and (8) which impose a set of restrictions across the parameters on the moving average (MA) representation. Given that the MA representation can be very closely approximated by a finite order VAR representation, Del Negro and Schorfheide (2004) propose to evaluate the DSGE model by assessing the validity of the restrictions imposed by such a model with respect to an unrestricted VAR representation. The choice of the variables to be included in the VAR is however completely driven by those entering in the DSGE model regardless of the statistical goodness of the unrestricted VAR. Policy variables set by optimization - typically included $\widetilde{\mathbf{Z}}_t$ - are naturally endogenous as optimal policy requires some response to current and expected developments of the economy. Expectations at time t for some of the variables of the systems at time $t+1$ are also included in the vector $\mathbf{Z}_t$,whenever the model is forward-looking. Models like (6) can be solved using standard numerical techniques as in Sims (2002) and the solution can be expressed as follows

$$\widetilde{\mathbf{Z}}_t = \mathbf{A}_0 + \mathbf{A}_1 \widetilde{\mathbf{Z}}_{\mathbf{t-1}} + \mathbf{R}\epsilon_t \tag{9}$$

where the matrices $\mathbf{A}_0, \mathbf{A}_1$,and $\mathbf{R}$ contain convolutions of the underlying model structural parameters. Consider the simple case in which all variables in the DSGE are observable and the number of structural shocks in ϵ_t is exactly equal to the number of variables in $\widetilde{\mathbf{Z}}_t$. In this case VAR are natural specifications for the data, therefore the estimated reduced form is

$$\widetilde{\mathbf{Z}}_t = \mathbf{A}_0 + \mathbf{A}_1 \widetilde{\mathbf{Z}}_{t-1} + u_t \tag{10}$$

Recent model evaluation of DSGE models exploits the fact that a solved RBC model is a statistical model. In fact, a solved DSGE model often generates a restricted MA representation for the vector of observable variables of interest, that can be approximated by a VAR of finite order (Fernandez-Villaverde *et al.,* 2007; Ravenna, 2007). Interestingly, this recent approach to model evaluation does not require identification of structural shocks but it is still potentially affected by lack of statistical identification. To make it clear, consider the general case of system (9) in which only a subset n of the m variables included in $\widetilde{\mathbf{Z}}_t$ is observable and define such a subset as Y_t. Now, Y_t has a VAR(∞) representation. This is usually approximated by a finite VAR representation at the cost of a truncation that can be relevant for purposes such as the identification of structural shocks (Ravenna 2007). Note that if the RBC model features a number of shocks smaller than the number of variables included in the VAR, some of the VAR shocks are interpreted as measurement error. The finite approximate VAR representation of a solved RBC model can be written taking into account the following system from Ravenna (2007)

$$\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + \mathbf{B}z_t \quad (11)$$
$$z_t = \mathbf{Z}_1 z_{t-1} + \varepsilon_t$$

where $\mathbf{Y}_t = \begin{bmatrix} x_t \\ y_t \end{bmatrix}$, x_t an $(n \times 1)$ vector of endogenous state variables, z_t an $(m \times 1)$ vector of exogenous state variables, y_t an $(r \times 1)$ vector of endogenous variables, ε_t an $(m \times 1)$ vector of stochastic process such that $E(\varepsilon_t) = 0, E(\varepsilon_t\varepsilon_t^{'}) = \Sigma, E(\varepsilon_t\varepsilon_\tau^{'}) = 0$ for $\tau \neq t$ and $\mathbf{\Sigma}$ is a diagonal matrix. All components of the vectors x_t and y_t are observable and the vector z_t has dimension $m = n + r$. Since the number of the observable variables, $n + r$, is equal to the number of the shocks, if $\mathbf{B}^{-1}$ exists, we can write a restricted VAR(2) representation of the system (11) as

$$\mathbf{Y}_t = (A + \mathbf{B}\mathbf{Z}_1\mathbf{B}^{-1})\mathbf{Y}_{t-1} - (\mathbf{B}\mathbf{Z}_1\mathbf{B}^{-1}\mathbf{A})\mathbf{Y}_{t-2} + \mathbf{B}\varepsilon_t$$

or

$$\mathbf{Y}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_1\mathbf{Y}_{t-1} + \mathbf{\Phi}_2\mathbf{Y}_{t-2} + u_t$$

where the VAR innovations $u_t = \mathbf{B}\varepsilon_t$ are a rotation of the structural shocks vector ε_t.

2.2 DSGE-VAR

The basic idea of the Del Negro-Schorfheide (2004) approach is to use the DSGE model to build prior distributions for the VAR. The starting point for the estimation is an unrestricted VAR of order p

$$\mathbf{Y}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_1\mathbf{Y}_{t-1} + ... + \mathbf{\Phi}_p\mathbf{Y}_{t-p} + \mathbf{u}_t \quad (12)$$

In compact format:

$$\mathbf{Y} = \mathbf{X}\mathbf{\Phi} + \mathbf{U} \quad (13)$$

$\mathbf{Y}$ is a $(T \times n)$ matrix with rows Y_t', $\mathbf{X}$ is a $(T \times k)$ matrix ($k = 1 + np, p =$number of lags) with rows $X_t' = [1, Y_{t-1}', ..., Y_{t-p}']$, $\mathbf{U}$ is a $(T \times n)$ matrix with rows u_t' and $\mathbf{\Phi}$ is a $(k \times n) = [\Phi_0, \Phi_1, ..., \Phi_p]'$.The one-step-ahead forecast errors u_t have a multivariate normal distribution $N(0, \mathbf{\Sigma}_u)$ conditional on past observations of Y.The log-likelihood function of the data is a function of $\mathbf{\Phi}$ and $\mathbf{\Sigma}_u$

$$L(\mathbf{Y}|\mathbf{\Phi}, \mathbf{\Sigma}_u) \propto |\mathbf{\Sigma}_u|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}tr\left[\Sigma_u^{-1}\left(\mathbf{Y}'\mathbf{Y} - \mathbf{\Phi}'\mathbf{X}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\mathbf{\Phi} + \mathbf{\Phi}'\mathbf{X}'\mathbf{X}\mathbf{\Phi}\right)\right]\right\} \quad (14)$$

The prior distribution for the VAR parameters proposed by Del Negro and Schorfheide (2004) is based on the statistical representation of the DSGE model given by a VAR approximation. Let Γ_{xx}^*, Γ_{yy}^*, Γ_{xy}^* and Γ_{yx}^* be the theoretical second-order moments of the variables $\mathbf{Y}$ and $\mathbf{X}$ implied by the DSGE model, where

$$\begin{aligned} \mathbf{\Phi}^*(\theta) &= \mathbf{\Gamma}_{xx}^{*-1}(\theta)\,\mathbf{\Gamma}_{xy}^*(\theta) \\ \mathbf{\Sigma}^*(\theta) &= \mathbf{\Gamma}_{yy}^*(\theta) - \mathbf{\Gamma}_{yx}^*(\theta)\,\mathbf{\Gamma}_{xx}^{*-1}(\theta)\,\mathbf{\Gamma}_{xy}^*(\theta) \end{aligned} \tag{15}$$

The moments are the dummy observation priors used in the mixture model. These vectors can be interpreted as the probability limits of the coefficients in a VAR estimated on the artificial observations generated by the DSGE model. Conditional on the vector of structural parameters in the DSGE model θ, the prior distributions for the VAR parameters $p(\mathbf{\Phi}, \mathbf{\Sigma}_u|\theta)$ are of the Inverted-Wishart (IW) and Normal forms

$$\begin{aligned} \mathbf{\Sigma}_u\,|\theta &\sim IW\left((\lambda T \mathbf{\Sigma}_u^*(\theta), \lambda T - k, n\right) \\ \mathbf{\Phi}\,|\mathbf{\Sigma}_u, \theta &\sim N\left(\mathbf{\Phi}^*(\theta), \mathbf{\Sigma}_u \otimes (\lambda T \mathbf{\Gamma}_{XX}(\theta))^{-1}\right) \end{aligned} \tag{16}$$

where the parameter λ controls the degree of model misspecification with respect to the VAR; for small values of λ the discrepancy between the VAR and the DSGE-VAR is large and a sizeable distance is generated between the unrestricted VAR and DSGE estimators. Large values of λ correspond to small model misspecification and for $\lambda = \infty$ beliefs about DSGE misspecification degenerate to a point mass at zero. Bayesian estimation could be interpreted as estimation based on a sample in which data are augmented by a hypothetical sample where observations are generated by the DSGE model, the so-called dummy prior observations (Theil and Goldberg 1961; Ingram and Whiteman 1994). Within this framework λ determines the length of the hypothetical sample.

The posterior distributions of the VAR parameters are also of the Inverted-Wishart and Normal forms. Given the prior distribution, posterior distributions are derived by the Bayes theorem

$$\mathbf{\Sigma}_u\,|\theta, \mathbf{Y} \sim IW\left((\lambda+1)\,T\hat{\mathbf{\Sigma}}_{u,b}(\theta), (\lambda+1)\,T - k, n\right) \tag{17}$$

$$\mathbf{\Phi}\,|\mathbf{\Sigma}_u, \theta, \mathbf{Y} \sim N\left(\hat{\mathbf{\Phi}}_b(\theta), \mathbf{\Sigma}_u \otimes [\lambda T \mathbf{\Gamma}_{XX}(\theta) + \mathbf{X}'\mathbf{X}]^{-1}\right) \tag{18}$$

$$\hat{\mathbf{\Phi}}_b(\theta) = (\lambda T \mathbf{\Gamma}_{XX}(\theta) + \mathbf{X}'\mathbf{X})^{-1}(\lambda T \mathbf{\Gamma}_{XY}(\theta) + \mathbf{X}'\mathbf{Y}) \tag{19}$$

$$\hat{\mathbf{\Sigma}}_{u,b}(\theta) = \frac{1}{(\lambda+1)\,T}\left[(\lambda T \mathbf{\Gamma}_{YY}(\theta) + \mathbf{Y}'\mathbf{Y}) - (\lambda T \mathbf{\Gamma}_{XY}(\theta) + \mathbf{X}'\mathbf{Y})\,\hat{\mathbf{\Phi}}_b(\theta)\right] \tag{20}$$

where the matrices $\hat{\mathbf{\Phi}}_b(\theta)$ and $\hat{\mathbf{\Sigma}}_{u,b}(\theta)$ have the interpretation of maximum likelihood estimates of the VAR parameters based on the combined sample of actual observations and artificial observations generated by the DSGE. Equations (17) and (18) show that the smaller λ is, the closer the estimates are to the OLS estimates of an unrestricted VAR. Instead, the higher λ is, the closer the VAR estimates will be tilted towards the parameters in the VAR approximation of the DSGE model ($\hat{\mathbf{\Phi}}_b(\theta)$ and $\hat{\mathbf{\Sigma}}_{u,b}(\theta)$). In order to obtain a non-degenerate prior density (16), which is a necessary condition for the existence of a well-defined Inverse-Wishart distribution, and for computing meaningful marginal likelihoods λ has to be greater than λ_{MIN}

$$\begin{aligned} \lambda_{MIN} &\geq \frac{n+k}{T}; k = 1 + p \times n \\ p &= \text{lags} \\ n &= \text{endogenous variables.} \end{aligned}$$

Hence, the optimal lambda must be greater than or equal to the minimum lambda $\left(\hat{\lambda} \geq \lambda_{MIN}\right)$.

Essentially, the DSGE-VAR tool allows the econometrician to draw posterior inferences about the DSGE model parameters θ. Del Negro and Schorfheide (2004) explain that the posterior estimate of θ has the interpretation of a minimum-distance estimator, where the discrepancy between the OLS estimates of the unrestricted VAR parameters and the VAR representation of the DSGE model is a sort of distance function. The estimated posterior of parameter vector θ depends on the hyperparameter λ. When $\lambda \to 0$, in the posterior the parameters are not informative, so the DSGE model is of no use in explaining the data. Unfortunately, the posteriors (18) and (17) do not have a closed form and we need a numerical method to solve the problem. The posterior simulator used by Del Negro and Schorfheide (2004) is the Markov Chain Monte Carlo Method and the algorithm used is the Metropolis-Hastings acceptance method. This procedure generates a Markov Chain from the posterior distribution of θ and this Markov Chain is used for Monte Carlo simulations. The optimal λ is given by maximizing the log of the marginal data density

$$\hat{\lambda} = \arg\max_{\lambda \geqslant \lambda_{MIN}} \ln p(\mathbf{Y}|\lambda)$$

According to the optimal lambda $\left(\hat{\lambda}\right)$, a corresponding optimal mixture model is chosen. This hybrid model is called DSGE-VAR$\left(\hat{\lambda}\right)$ and $\hat{\lambda}$ is the weight of the priors. It can also be interpreted as the restriction of the theoretical model on the actual data.

2.3 Other Models

In order to evaluate the forecasting performance of the simple DSGE model and the DSGE-VAR, the classical VAR as well as the Bayesian VAR are also implemented.

2.3.1 Classical VAR

The classical unrestricted VAR, as suggested by Sims (1980), has the following compact format

$$\mathbf{Y}_t = \mathbf{X}_t \mathbf{\Phi} + \mathbf{U} \tag{21}$$

where $\mathbf{Y}_t$ is a $(T \times n)$ matrix with rows Y_t', $\mathbf{X}$ is a $(T \times k)$ matrix ($k = 1 + np, p =$number of lags) with rows $X_t' = [1, Y_{t-1}', ..., Y_{t-p}']$. $\mathbf{U}$ is a $(T \times n)$ matrix with rows u_t', $\mathbf{\Phi}$ is a $(k \times n) = [\Phi_0, \Phi_1, ..., \Phi_p]'$, while the one-step ahead forecast errors u_t have a multivariate $N(0, \mathbf{\Sigma}_u)$ conditional on past observations of $\mathbf{Y}$.

2.3.2 Bayesian VAR

The Bayesian VAR, as described in Litterman (1981), Doan *et al.* (1984), Todd (1984), Litterman (1986) and Spencer (1993) has become a widely popular approach to dealing with overparameterization. One of main problems in using VAR models is that many parameters need to be estimated, although some of them may be insignificant. Instead of eliminating longer lags, the BVAR imposes restrictions on these coefficients by assuming that they are more likely to be near zero than the coefficients on shorter lags. Obviously, if there are strong effects from less important variables, the data can counter this assumption. Usually, the restrictions are imposed by specifying normal prior distributions with zero means and small standard deviations for all coefficients, with a decreasing standard deviation as the lags increase. The only exception is the coefficient on a variable's first lag that has a mean of unity. Litterman (1981) used a diffuse prior for the constant. The means of the prior are popularly called the "Minnesota Priors" due to the development of the idea at the University of Minnesota and the Federal Reserve Bank at Minneapolis.

Formally speaking, these prior means can be written as follows

$$\Phi_i \sim \begin{cases} (\delta_i, \sigma^2_{\Phi_i}), & j = i, k = 1 \\ (0, \sigma^2_{\Phi_j}), & \text{otherwise} \end{cases} \tag{22}$$

where Φ_i denotes the coefficients associated with the lagged dependent variables in each equation of the VAR, while Φ_j represents any other coefficient. Litterman's prior was designed for data in levels and has the effect of shrinking the process towards the univariate random walk. Consequently, δ_i is equal 1 for all i, reflecting the high persistence. In this study, we impose their prior mean on the first own lag for variables in growth rate, such as a white noise setting $\delta_i = 0$ (Del Negro and Schorfheide 2004; Adolfson *et al.* 2007; Banbura *et al.* 2010). Instead, for level variables, we use the classical Minnesota prior (Del Negro and Schorfheide 2004).

Doan *et al.* (1984) propose a formula to generate standard deviations as a function of a small number of hyperparameters w and d, and a weighting matrix $F(i, j)$. This approach is useful for the forecaster to specify individual prior variances for a large number of coefficients based on only a few hyperparameters. The specification of the standard deviation of the distribution of the prior imposed on variable j in equation i at lag m, for all i, j and m, denoted by $S(i, j, m)$, is specified as follows

$$S(i, j, m) = [w \times g(m) \times F(i, j)] \frac{\hat{\sigma}_i}{\hat{\sigma}_j}, \tag{23}$$

where

$$F(i, j) = \begin{cases} 1 \text{ if } i = j \\ k_{ij} \text{ otherwise, } 0 \leq k_{ij} \leq 1 \end{cases} \tag{24}$$

is the tightness of variable j in equation i relative to variable i and by increasing the interaction, i.e. it is possible for the value of k_{ij} to loosen the prior (Dua and Ray 1995). Reducing the interaction parameter k_{ij} tightens the prior. The ratio $\frac{\hat{\sigma}_i}{\hat{\sigma}_j}$ consists of estimated standard errors of the univariate autoregression, for variables i and j. This ratio scales the variables to account for differences in the units of measurement, without taking into account the magnitude of the variables. The term w measures the standard deviation on the first lag, and also indicates the overall tightness of the prior distribution around the random walk or white noise. The overall tightness governs the relative importance of the prior beliefs with respect to the information contained in the data. A decrease in the value of w results in a tighter prior. For $w = 0$ the posterior equals the prior and the data do not influence the estimates. If $w \to \infty$, the posterior expectations coincide with the ordinary least squares estimates. As pointed in Banbura *et al.* (2010), for small and medium VAR models, the choice of the overall tightness matters. The selection of the shrinkage coefficient w can be done subjectively as in Litterman (1986), or we can choose it in relation to the size of the system. As the number of variables increases, the parameters should be shrunk more to avoid overfitting (De Mol *et al.* 2008). The selection can be implemented via a grid search in a training sample as in Banbura *et al.* (2010). The function $g(m) = m^{-d}$, $d > 0$ is the measurement of the tightness on lag m relative to lag 1, and is assumed to have a harmonic shape with a decay of d, which tightens the prior on increasing lags. Obviously, the function $g(m)$ is different across lags.

3 State space time-varying parameter VAR model

The classical VAR modelling is adequate only in the analysis of stationary time series, and in many cases, stationarity assumptions are too restrictive. In these cases, the use of time-varying parameters seems to be an attractive alternatives. Time varying autoregression (TVP-VAR) models have been developed since the early 1980's. Primiceri (2005) used them in analyzing macroeconomic policy issues. The TVP-VAR model enables capturing a possible time-varying nature of underlying structure in the economy in a robust manner by allowing both temporary and permanent shift in the parameters.

In this paper, we propose a novel time-varying multivariate state-space estimation method for VAR models. Regarding the parameters of the TVP-VAR as state parameter variables, TVP autoregression could easily form a state space model. The state space model has been well studied by Harvey (1990) and Durbin and Koopman (2002). According to Kalman (1960, 1963), in a state-space representation the signal extraction is implemented through a model that links the unobserved and observed variables of the system. To estimate a state space model, several methods have been developed. Kalman filtering involves sequentially updating a linear projection on the vector of interest. The state-space representation is given by a system of two vector equations. First, the state or transition equation describes the dynamics of the state vector containing the unobserved variables we estimate, while the second equation represents the observation or measurement equation linking the state vector to the vector containing the observed variables. For the TVP-VAR models, the parameters are estimated using a multivariate specification of the standard Kalman filter (Harvey 1990) . The likelihood estimation requires repeating the filtering many times in order to evaluate the likelihood for each set of the time-varying parameters until we reach the maximum. This is performed with a suitable multivariate extension of the Kim and Nelson (1999) method. The calculation of the Hessian for the estimation of the variance-covariance matrix is done with the Broyden-Fletcher-Goldfarb-Shano (BFGS) optimization algorithm. Other algorithms can also be used with the same results, e.g., the DFP and the Levenberg-Marquardt. The parameters could be also estimated with the use of the Zellner g-prior and in this case the numerical evaluation of the posterior distributions is performed with Gibbs sampling (Kim and Nelson 1999).

The TVP-VAR can be expressed as

$$\mathbf{y}_t = \mathbf{\Phi}_{0,t} + \mathbf{\Phi}_{1,t}\mathbf{y}_{t-1} + \cdots + \mathbf{\Phi}_{p,t}\mathbf{y}_{t-p} + u_t \tag{25}$$

in which $\mathbf{\Phi}_{0,t}$ is a $k \times 1$ vector of time-varying intercepts, $\mathbf{\Phi}_{i,t}$ $(i = 1, \ldots, p)$ are $k \times k$ matrices of time-varying coefficients and u_t are homoscedastic or heteroscedastic reduced-form residuals with a covariance matrix $\mathbf{\Omega}_t$. This could be transformed into a multivariate state-space form.

Harvey (1990) provides a framework for a multivariate version of the Kalman filter based on a time series analogue of the seemingly unrelated regression equation (SURE) model introduced into econometrics by Zellner (1963). Harvey (1990) refers to it as a system of seemingly unrelated time series equations (SUTSE) model. An important property of the SUTSE system is that its form remains unaltered when it is subject to contemporaneous aggregation. A linear time-invariant univariate structural model can be written in the SUTSE state space form for N variables

$$\mathbf{Y}_t = (\mathbf{z}' \otimes \mathbf{I}_N)\, \boldsymbol{\alpha}_t + \varepsilon_t \tag{26}$$

$$\boldsymbol{\alpha}_t = (\mathbf{T} \otimes \mathbf{I}_N)\, \boldsymbol{\alpha}_{t-1} + (\mathbf{R} \otimes \mathbf{I}_N)\, \boldsymbol{\eta}_t \tag{27}$$

where $\boldsymbol{\alpha}_t$ is the vector of state variables of the the *state* equation, $\mathbf{T}_t$ the state transition matrix, $Var(\boldsymbol{\varepsilon}_t) = \boldsymbol{\Sigma}_\varepsilon$ and $Var(\boldsymbol{\eta}_t)$ a block diagonal matrix with the blocks being $\boldsymbol{\Sigma}_k$, $k = 1, ..., g$. For example, in the three-variate case, the variance of the error component in the state equation is

$$Var(\boldsymbol{\eta}_t) = \begin{bmatrix} \boldsymbol{\Sigma}_\eta & 0 & 0 \\ 0 & \boldsymbol{\Sigma}_\zeta & 0 \\ 0 & 0 & \boldsymbol{\Sigma}_\omega \end{bmatrix} \tag{28}$$

In fact a more general formulation of the SUTSE model does not constrain $Var(\boldsymbol{\eta}_t)$ to be diagonal and hence $Var(\boldsymbol{\eta}_t)$ need not be block diagonal. Indeed the SUTSE formulation can be generalized further to allow quantities such as $\mathbf{z}$, $\boldsymbol{\Sigma}_\varepsilon, \mathbf{T}, \mathbf{R}$ and $Var(\boldsymbol{\eta}_t)$ to change deterministically over time. As shown in Harvey (1986), the time-domain treatment still goes through. The Kalman filter may be applied to (26) and (27), the number of sets of observations needed to form an estimator of $\boldsymbol{\alpha}_t$, with finite MSE matrix being the same as in the univariate case. The conditions for the filter to converge to a steady state are an obvious generalization of the conditions in the univariate case. Given normality of the disturbances, the log-likelihood function is of the prediction error decomposition form.

The decoupling of the Kalman filter is related to the result which arises in a SURE system where OLS applied to each equation in turn is fully efficient if each equation contains the same regressors. Hence, all the information needed for estimation, prediction and smoothing can be obtained by applying the same univariate filter to each series in turn. Consider the multivariate random walk plus noise model. If the signal-to-noise ratio is q (i.e., $\boldsymbol{\Sigma}_\eta / \boldsymbol{\Sigma}_\varepsilon = q$), the Kalman filter for this model is

$$\boldsymbol{\alpha}_{t+1|t} = \boldsymbol{\alpha}_{t|t-1} + \mathbf{K}_t \left(\mathbf{Y}_t - \boldsymbol{\alpha}_{t|t-1} \right), t = 2, \ldots, T \tag{29}$$

and

$$\mathbf{P}_{t+1|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{F}_t^{-1} \mathbf{P}_{t|t-1} + q \boldsymbol{\Sigma}_\varepsilon \tag{30}$$

where

$$\mathbf{K}_t = \mathbf{P}_{t|t-1} \mathbf{F}_t^{-1} \tag{31}$$

and

$$\mathbf{F}_t = \mathbf{P}_{t|t-1} + \boldsymbol{\Sigma}_\varepsilon \tag{32}$$

Let w_t denote a positive scalar for $t = 2, \ldots, T$ and suppose that $\mathbf{P}_{t|t-1}$, the MSE matrix of the $N \times 1$ vector $\boldsymbol{\alpha}_{t|t-1}$, is proportional to $\boldsymbol{\Sigma}_\varepsilon$, i.e. $\mathbf{P}_{t|t-1} = w_t \boldsymbol{\Sigma}_\varepsilon$. It then follows from (30) that $\mathbf{P}_{t+1|t}$ is of the same form, that is, $\mathbf{P}_{t+1|t} = w_{t+1} \boldsymbol{\Sigma}_\varepsilon$ with $w_{t+1} = (w_t + w_t q + q) / (w_t + 1)$. Furthermore if $\mathbf{P}_{t|t-1} = w_t \boldsymbol{\Sigma}_\varepsilon$ the gain matrix in (29) is diagonal, that is

$$\mathbf{K}_t = w_t \boldsymbol{\Sigma}_\varepsilon \left(w_t \boldsymbol{\Sigma}_\varepsilon + \boldsymbol{\Sigma}_\varepsilon \right)^{-1} = \left[w_t / \left(w_{t+1} \right) \right] \mathbf{I}_N \tag{33}$$

Suppose that the above Kalman filter is started off in such a way that $\mathbf{P}_{2|1}$ is proportional to $\boldsymbol{\Sigma}_{\varepsilon}$; that is $\mathbf{P}_{2|1} = p_{_{2|1}} \boldsymbol{\Sigma}_{\varepsilon}$, where $p_{_{2|1}}$ is a scalar. Since $\mathbf{P}_{t|t-1}$ must continue to be proportional to $\boldsymbol{\Sigma}_{\varepsilon}$, it follows from (33) that the elements of $\boldsymbol{\alpha}_{t+1|t}$, can be computed from the univariate recursions. It also follows that w_t, must be equal to $p_{_{t|t-1}}$ for all $t = 2, \ldots, T$. The starting values $\boldsymbol{\alpha}_{_{2|1}} = \mathbf{y}_1$ and $\mathbf{P}_{2|1} = \boldsymbol{\Sigma}_{\eta} + \boldsymbol{\Sigma}_{\varepsilon} = (1+q)\,\boldsymbol{\Sigma}_{\varepsilon}$ equally correspond to the use of a diffuse prior, and the use of these starting values leads to the exact likelihood function for $\mathbf{Y}_2, \ldots, \mathbf{Y}_T$ in the prediction error decomposition form

$$\log L = -\frac{(T-1)\,N}{2}\log 2\pi - \frac{1}{2}\sum_{t=2}^{T}\log|\mathbf{F}_t| - \frac{1}{2}\sum_{t=2}^{T}\mathbf{v}_t'\mathbf{F}_t^{-1}\mathbf{v}_t \qquad (34)$$

However, the decoupling of the Kalman filter allows the elements of $\mathbf{v}_t$, to be computed from the univariate recursions. Furthermore

$$\mathbf{P}_{t|t-1} = p_{_{t|t-1}}\boldsymbol{\Sigma}_{\varepsilon} \qquad (35)$$

and so

$$\mathbf{F}_t = \mathbf{P}_{t|t-1} + \boldsymbol{\Sigma}_{\varepsilon} = f_t\boldsymbol{\Sigma}_{\varepsilon},\ t = 3, \ldots, T \qquad (36)$$

where $f_t = \left(p_{_{t|t-1}} + 1\right)$. Substituting from (36) into (34) gives

$$\log L = -\frac{(T-1)\,N}{2}\log 2\pi + \frac{(T-1)}{2}\log\left|\boldsymbol{\Sigma}_{\varepsilon}^{-1}\right| - \frac{N}{2}\sum_{t=2}^{T}\log f_t - \frac{1}{2}\sum_{t=2}^{T}\frac{1}{f_t}\mathbf{v}_t'\boldsymbol{\Sigma}_{\varepsilon}^{-1}\mathbf{v}_t \qquad (37)$$

Differentiating (37) with respect to the distinct elements of $\boldsymbol{\Sigma}_{\varepsilon}^{-1}$ leads to the ML estimator of $\boldsymbol{\Sigma}_{\varepsilon}$ being

$$\tilde{\boldsymbol{\Sigma}}_{\varepsilon} = (T-1)^{-1}\sum_{t=2}^{T} f_t^{-1}\mathbf{v}_t\mathbf{v}_t' \qquad (38)$$

for any given value of q. The ML estimators of q and $\boldsymbol{\Sigma}_{\varepsilon}$ can therefore be obtained by maximizing the concentrated likelihood function

$$\log L_c = -\frac{(T-1)\,N}{2}\log 2\pi - \frac{(T-1)}{2}\log\left|\tilde{\boldsymbol{\Sigma}}_{\varepsilon}\right| - \frac{N}{2}\sum_{t=2}^{T}\log f_t \qquad (39)$$

with respect to q. Once the parameters have been estimated, prediction and smoothing can be carried out. The predictions of future observations are obtained from the univariate recursions

$$\mathrm{MSE}\left(\tilde{\mathbf{y}}_{T+l|T}\right) = f_{T+l|T}\boldsymbol{\Sigma}_{\varepsilon},\ l = 1, 2, \ldots \qquad (40)$$

where

$$f_{T+l|T} = p_{T+l|T} + 1 \tag{41}$$

The decoupling of the Kalman filter can be shown in a similar way for the time-varying system

$$\mathbf{Y}_t = (\mathbf{z}_t' \otimes \mathbf{I}_N)\,\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, Var\,(\boldsymbol{\varepsilon}_t) = h_t \boldsymbol{\Sigma}_* \tag{42}$$

$$\boldsymbol{\alpha}_t = (\mathbf{T}_t \otimes \mathbf{I}_N)\,\boldsymbol{\alpha}_{t-1} + (\mathbf{R}_t \otimes \mathbf{I}_N)\,\boldsymbol{\eta}_t,\ Var\,(\boldsymbol{\eta}_t) = \mathbf{Q}_t \otimes \boldsymbol{\Sigma}_* \tag{43}$$

where $\mathbf{Q}_t = \text{diag}(q_1, \ldots, q_k)$. The more general formulation does not constrain $\mathbf{Q}_t$ to be diagonal, although, as in the univariate model, restrictions are needed on $\mathbf{Q}_t$ for the model to be identifiable. All the results on estimation and prediction carry through, with $\mathbf{P}_{t+1|t} = \mathbf{P}^*_{t+1|t} \otimes \boldsymbol{\Sigma}_*$, where $\mathbf{P}^*_{t+1|t}$ is the MSE matrix for the univariate model (Harvey 1986, 1990).

4 Empirical results

The models are estimated based on quarterly data of the US economy over the total period 1980:1 to 2009:4 following the works of Benati and Surico (2008), Schorfheide *et al.* (2010), Cogley *et al.* (2008), Herbst and Schorfheide (2012) and Consolo *et al.* (2009). The sample begins in 1980 when signifying changes in US monetary and fiscal policy occurred (Ireland 2004) and thus it can be considered a major breakpoint. The starting period roughly coincides with the end of the Volcker stabilization and disinflation era. We concentrate on the 1980-2009 period which is characterized by a more stable monetary and financial structure and a lower volatility of the macroeconomic variables. Structural breaks in mean and volatility are found in the literature by comparing the pre-80 with the post-80 period, while the null hypothesis of parameter stability cannot be rejected in the post-80 period (Justiniano and Primiceri 2008). Moreover inflation, monetary policy rate, annual real output growth and other variables used in the empirical literature are clearly mean reverting in the post 1980 period. Consolo *et al.* (2009) mention that this evidence reduces the concern of having a non-stationary VAR that omits potential long-run cointegrating relations among the variables of interest. Furthermore, Benati and Surico (2008) claim that if the U.S. economy was indeed in an indeterminate equilibrium before but not after October 1979, then by estimating TVP-VAR and DSGE models before and after the 80s they would be mixing two quite different regimes, thus obtaining biased estimates of the structural parameters. Finally, Herbst and Schorfheide (2012) argue that since there is strong empirical evidence that monetary policy as well as the volatility of macroeconomic shocks changed in the early 1980s, the information set in estimating DSGE and TVP-VAR models should be relevant to the exercise of contemporary policy making, thus a sample after the 1980s ensures a better forecasting performance.

The data for real output growth comes from the Bureau of Economic Analysis as Gross Domestic Product (GDP-SAAR, billions chained 2005). Consumer price index data are derived from the Bureau of Labor Statistics (CPI-U: all items, seasonally adjusted, 1982-1984=100). GDP and CPI are taken in first difference logarithmic transformation. The interest rate series (FFR) are constructed as in Clarida *et al.* (2000); for each quarter the interest rate is computed as the average federal funds rate (source: Haver Analytics) during the first month of the quarter, including business days only. These three time series represent the three equations of the DSGE and VAR model classes. We compare the out-of-sample forecasting performance of VAR, Bayesian VAR, DSGE and DSGE-VAR models and the multivariate state space TVP-VAR, in terms of the Root Mean Squared Error (RMSE) for different lag specifications (one to four). The out-of-sample period is 2001:1-2009:4. The forecasting investigation is performed over the one- to five-quarter-ahead horizon with a rolling estimation sample, based on the works of Marcellino (2004) and Brüggemann *et al.* (2008) for datasets of quarterly frequency. In particular, the models are re-estimated each quarter over the forecast horizon to update the estimate of the coefficients, before producing the one- to five-quarters-ahead forecasts. Then, in order to evaluate the models' forecast accuracy, we use the cross-model test statistic of Diebold and Mariano (1995) and the Clark and West (2004) test for nested (restricted) models which is based on Clark and McCracken (2001) and Newey-West estimator (1987, 1994) of the asymptotic variance matrix. The application of the Clark and West (2004) test for comparing the out-of-sample accuracy of two models was considered necessary as many of the competing models are nested and this causes the mean prediction error of the restricted model to be often smaller than that of the alternative, leading to size and power distortions of the Diebold and Mariano test. The larger the number of parameters in the unrestricted model the larger the difference will be. In addition, the Clark-West test uses the Newey-West estimator (1987, 1994) to correct for the autocorrelation of the forecast errors, since the Diebold-Mariano assumes for h-steps-ahead forecasts all autocorrelations of order equal or greater to h for the squared forecast errors difference are zero and consequently the empirical autocorrelation of the errors tends to be of higher order than h.

Regarding the BVAR setup, in our empirical exercise we impose the prior mean on the first own lag for GDP and CPI as a white noise setting $\delta_i = 0$ (Adolfson, *et al.* 2007; Del Negro and Schorfheide 2004; Banbura *et al.* 2010). Instead, for the FFR interest rate, we use the classical Minnesota prior (Del Negro and Schorfheide 2004). For the selection of the hyperparameters, we follow the empirical strategy proposed by Liu *et al.* (2009) and Gupta and Kabundi (2010). We implement a grid search to optimize the forecasting performance. In case of overall tightness, we consider the grid $(0.1, 0.2, 0.3)$, and for the lag decay, a grid search $(0.5, 1, 2)$. For the interaction term, instead of considering a symmetric interaction function $F(i, j)$, assuming $k_{ij} = 0.5$ as in Dua and Smyth (1995) we check different values $(0.1, 0.5, 0.9)$, to allow for a different tightness for the interactions. Eventually, we optimize forecasting across different lags for the BVAR with $w = 0.1$, $k_{ij} = 0.1, d = 0.5$.

The prior distribution for the DSGE model parameters (θ), which are similar to the priors used by Del Negro and Schorfheide (2004), are illustrated in Table 1. Using these priors, we can build the hybrid models, taking into account that the statistical representation of the DSGE is given by a restricted VAR(2). In the forecasting exercise, the forecasted values are produced implementing the DSGE-VAR ($\hat{\lambda}$), where the $\hat{\lambda}$ is chosen by the numerical procedure for each estimation. In the forecasting evaluation, the DSGE-VAR is estimated with a different number of lags on the sample spanning from 1980:1 to 2000:4. As already mentioned the out-of-sample forecasting accuracy is assessed based on a rolling sample and the DSGE-VARs are re-estimated for each rolling sample. The forecasts are calculated as a mean of forecast draws, taking into account the $\hat{\lambda}$ found in the in-sample estimation. Parameter λ is chosen from a grid which is unbounded from above. In our empirical exercise, the log of the marginal data density is computed over a discrete interval, $\ln p(Y|\lambda, M)$. The minimum value, $\lambda_{\min} = \frac{n+k}{T}$, is model dependent and is related to the existence of a well-defined Inverse-Wishart distribution. For completeness, it is worth mentioning that $\lambda = 0$ refers to the VAR model with no prior and it is not possible to compute the marginal likelihood in this particular case. Therefore, we can show the log of marginal data density for any value of λ larger than $\lambda_{\min}$. Thus, $\lambda_{\min}$ depends on the degrees of freedom in the VAR.[1]

Table 1 Prior Distributions for the DSGE model parameters

Name	Range	Density	Starting value	Mean	Standard deviation
$\ln \gamma$	$\mathbb{R}$	Normal	0.500	0.500	0.250
$\ln \pi^*$	$\mathbb{R}$	Normal	1.000	1.000	0.500
$\ln r^*$	$\mathbb{R}^+$	Gamma	0.500	0.500	0.250
κ	$\mathbb{R}^+$	Gamma	0.400	0.300	0.150
τ	$\mathbb{R}^+$	Gamma	1.000	2.000	0.500
ψ_1	$\mathbb{R}^+$	Gamma	2.500	1.500	0.250
ψ_2	$\mathbb{R}^+$	Gamma	0.300	0.125	0.100
ρ_R	$[0, 1)$	Beta	0.400	0.500	0.200
ρ_G	$[0, 1)$	Beta	0.800	0.800	0.100
ρ_Z	$[0, 1)$	Beta	0.200	0.300	0.100
σ_R	$\mathbb{R}^+$	Inv.Gamma	0.500	0.251	0.139
σ_G	$\mathbb{R}^+$	Inv.Gamma	0.500	0.630	0.323
σ_Z	$\mathbb{R}^+$	Inv.Gamma	1.000	0.875	0.430

Note: The model parameters $\ln \gamma$, $ln\pi^*$, $\ln r^*$, σ_R, σ_g, and σ_z are scaled by 100 to convert them into percentages. The Inverse Gamma priors are of the form $p(\sigma|\nu, s) \propto \sigma^{-\nu-1} e^{-\nu s^2/2\sigma^2}$, where v=4 and s equals 0.2, 0.5, and 0.7, respectively. Approximately 1.5% of the prior mass lies in the indeterminacy region of the parameter space. The prior is truncated to restrict it to the determinacy region of the DSGE model.

[1] For the DSGE-VAR, the lambda grid is given by

$$\Lambda = \left\{ \begin{array}{c} 0,\ 0.06,\ 0.09,\ 0.12,\ 0.14,\ 0.15,\ 0.20,\ 0.25,\ 0.30,\ 0{,}35,\ 0.40, \\ 0.45,\ 0.50,\ 0.55,\ 0.60,\ 0.65,\ 0.70,\ 0.75,\ 0.80,\ 0.85,\ 0.9,\ 1,\ 5,\ 10 \end{array} \right\}.$$

In this lambda interval, we consider the λ_{MIN} across lags from one to four.

Table 2 Optimal lambda for the DSGE-VAR calculated with Markov Chain Monte Carlo and Metropolis Hastings method

	λ_{MIN}	$\hat{\lambda}$	$\hat{\lambda}-\lambda_{MIN}$	$\frac{\hat{\lambda}-\lambda_{MIN}}{\lambda_{MIN}}$	$\ln p(Y\|\hat{\lambda}, M)$	Bayes Factor
DSGE-VAR(1)	0.06	0.09	0.03	0.5	-486.5	$exp\,[26.3]$
DSGE-VAR(2)	0.09	0.70	0.61	6.8	-460.2	1
DSGE-VAR(3)	0.12	0.30	0.18	1.5	-450.8	$exp\,[-9.4]$
DSGE-VAR(4)	0.14	0.70	0.56	4.0	-414.3	$exp\,[-45.9]$

Table 2 shows the main results related to the DSGE-VAR implemented using a different number of lags (from one up to four) in case of one-step ahead[2]. Each minimum λ (λ_{MIN}) is given by the features of the model (number of observations, number of endogenous variables, number of lags) and the optimal lambda ($\hat{\lambda}$) is calculated using the Markov Chain Monte Carlo with Metropolis Hastings acceptance method and 10,000 replications for each of 10 MH blocks. All Markov Chain Monte Carlo results are based on 110,000 draws from the relevant posterior distribution, discarding the first 10,000. We checked whether 110,000 draws were sufficient by repeating the MCMC computations from overdispersed starting points, and we verified that we obtained the same results for parameter estimates and log-marginal likelihood functions. The $\ln p(Y|M)$ is the log marginal data density for the DSGE model specifications computed based on Geweke's (1999) modified harmonic mean estimator. The Bayes factor (ratio of posterior odds to prior odds) (Schorfheide 2010) helps us to understand the improvement of the log marginal data density of a specific model compared to a benchmark model, which for the MCMC exercise is the DSGE-VAR (2), since the statistical representation of the DSGE model is given by a VAR(2). According to Table 2, the difference $\hat{\lambda} - \lambda_{MIN}$ is the greatest in the case of a DSGE-VAR(2), and hence its corresponding ratio $\frac{\hat{\lambda}-\lambda_{MIN}}{\lambda_{MIN}}$ is the greatest too. Looking at the log of the marginal data densities, we notice that the DSGE-VAR(4) model has the minimum value and the Bayes factor evidences a great difference between the DSGE-VAR(2) (the benchmark model) and the DSGE-VAR(4) ($exp\,[-45.9]$) in favour of the DSGE-VAR with four lags.

Tables 3, 4 and 5 report the RMSE ratios against the benchmark model for the forecasting exercise, which in this study is VAR(2). The number of lags was indicated by the Schwartz Bayesian information criterion (SIC) criterion

[2] Table 2 presents the $\widehat{\lambda}$ calculated in one-step ahead exercise. However, the optimal lambda for each lag length is very similar across forecasting samples. For further steps ahead results are upon request.

Table 3 RMSE Ratios for the US GDP (total period)

GDP	Quarters ahead				
	1	2	3	4	5
VAR(1)	0.970	0.960	0.973	0.953	0.939
VAR(3)	0.980	0.981	0.980	0.971	0.988
VAR(4)	0.961	0.973	0.953	0.956	0.970
BVAR(1)	0.959	0.937	0.976	0.940	0.936
BVAR(2)	1.002	0.983	1.023	0.995	1.001
BVAR(3)	0.980	0.959	0.997	0.972	0.992
BVAR(4)	0.955	0.938	0.977	0.944	0.956
DSGE	1.055	1.154	1.083	1.181	1.278
DSGE-VAR(1)	0.959	0.965	0.945	0.928	0.874
DSGE-VAR(2)	0.991	1.012	0.942	0.857	0.875
DSGE-VAR(3)	0.994	1.017	0.922	0.906	0.913
DSGE-VAR(4)	1.027	1.012	0.888	0.883	0.897
TVP-VAR(1)	0.960	0.858	0.892	0.868	0.817
TVP-VAR(2)	0.965	0.799	0.873	0.854	0.822
TVP-VAR(3)	1.007	0.838	0.896	0.844	0.806
TVP-VAR(4)	1.017	0.895	0.893	0.852	0.789

Notes: The ratios are estimated against the benchmark model VAR(2) for one- to five-steps-ahead in the out-of-sample (rolling) period 2001:1 to 2009:4.

Table 4 RMSE Ratios for the US CPI (total period)

CPI	Quarters ahead				
	1	2	3	4	5
VAR(1)	1.006	0.998	0.986	0.996	0.998
VAR(3)	0.988	0.986	0.994	0.995	0.996
VAR(4)	0.991	0.983	0.982	0.994	0.989
BVAR(1)	0.998	1.004	1.002	0.994	1.004
BVAR(2)	1.001	1.009	1.008	0.996	1.004
BVAR(3)	0.987	0.998	1.004	0.992	1.013
BVAR(4)	0.989	0.995	0.996	0.990	1.004
DSGE	0.996	1.002	0.974	0.990	1.028
DSGE-VAR(1)	0.993	1.000	0.975	0.990	1.009
DSGE-VAR(2)	1.000	0.991	0.980	0.990	1.026
DSGE-VAR(3)	0.976	0.979	0.981	0.995	1.066
DSGE-VAR(4)	0.983	0.979	0.984	1.002	1.041
TVP-VAR(1)	1.070	1.013	1.014	1.032	1.001
TVP-VAR(2)	1.186	0.960	1.030	1.037	1.012
TVP-VAR(3)	1.188	1.087	1.082	1.114	1.106
TVP-VAR(4)	1.155	1.060	1.066	1.053	0.988

Notes: As in Table 3

for the simple vector autoregression system. For the GDP series the TVP-VAR with a multivariate state-space representation outperforms the other models for all the quarters-ahead forecasts except for the one-quarter-ahead where the BVAR(4) provides a better RMSE ratio against the benchmark. The TVP-VAR achieves a better score for the RMSE ratios with three lags for the two-quarter and four-quarters ahead forecasts, and two and four lags for the three-quarter and five-quarter ahead forecasts respectively. The DSGE-VAR is in general better than the simple DSGE which on average generates the

Table 5 RMSE Ratios for the US FFR (total period)

FFR	Quarters ahead				
	1	2	3	4	5
VAR(1)	1.075	1.082	0.996	1.114	1.349
VAR(3)	0.865	0.838	0.913	0.854	0.853
VAR(4)	0.870	0.797	0.783	0.822	0.980
BVAR(1)	1.030	1.019	0.920	1.009	1.186
BVAR(2)	1.019	1.003	0.884	0.708	0.889
BVAR(3)	0.931	0.917	0.839	0.768	0.747
BVAR(4)	0.944	0.894	0.815	0.864	1.046
DSGE	0.890	0.968	0.931	1.027	1.399
DSGE-VAR(1)	1.068	1.201	0.681	0.730	0.709
DSGE-VAR(2)	0.963	1.043	0.712	0.562	0.730
DSGE-VAR(3)	0.908	0.978	0.700	0.568	0.749
DSGE-VAR(4)	0.931	0.923	0.646	0.548	0.772
TVP-VAR(1)	0.420	0.429	0.484	0.640	0.918
TVP-VAR(2)	0.447	0.460	0.498	0.647	0.919
TVP-VAR(3)	0.510	0.536	0.560	0.689	0.946
TVP-VAR(4)	0.502	0.530	0.548	0.664	0.889

Notes: As in Table 3

worst forecast ratios, while VARs and BVARs present similar predictive performance, albeit BVAR slightly better for all steps-ahead. The DSGE-VAR with any lag structure seems to be better than VAR and BVARs particularly over the three-quarter ahead forecasts of the GDP The results for the CPI series vary. Specifically, the DSGE-VAR(3) provides the lowest RMSE ratio against the benchmark for the one-quarter-ahead forecast, while the state-space TVP-VAR outperforms the other models for the two- and five-quarter-ahead forecasts. Interestingly, the results suggest that the simple DGSE produces lowest ratios for three- and four-steps ahead, although this is compensated by an equal ratio by the DSGE-VAR(1), DSGE-VAR(2) and BVAR(4) for the four-steps ahead forecasts. Overall, the TVP-VAR underperforms relatively to the other models, while on average VARs provide better results than the other specifications. Regarding the FFR variable, the TVP-VAR and the DSGE-VAR models are the winners. In particular, for the first three quarter-ahead forecasts the TVP-VAR(1) clearly outperforms all other models, while for the four- and five-quarter-ahead the DSGE-VAR(4) and DSGE-VAR(1) respectively give the best results against the benchmark. Generally, VAR and BVAR models provide with similar ratios while on average the DSGE-VAR is better than the simple DSGE[3].

[3] Via the utilization of a disjoint larger sample and beyond the current dataset, we performed a sensitivity analysis with respect to the initial sample selection. This could also be considered as a further robustness check regarding the forecast error results as well as the differential predictability results presented later. As the current database did not include a longer sample with the exact same variables/time series we used a new sample that spanned from 1955:1 to 2009:4 with an out-of-sample period of 2001:1-2009:4 (same as the initial exercise). The real Gross Domestic Product (GDP) and the Consumer Price Index (CPI) came from the Historical Data Files for the Real-Time Data Set provided by the Federal Reserve Bank of Philadelphia. The short interest rate series came from the ALFRED dataset for vintage data provided by the Federal Reserve Bank of St. Louis. The quarterly real GDP

Furthermore, we discuss results concerning in particular the US financial crisis period 2008 -2009 in order to juxtapose them against that of the total forecasting period. We consider the out-of-sample sub-period 2008:1-2009:4 and again we report the RMSE for the various models against the benchmark VAR(2) over the one- to five-quarter-ahead horizon with a rolling estimation sample. In case of GDP, the TVP-VAR specifications still achieve better scores for the RMSE ratios for two- and three-quarters ahead forecasts as in the total period, while it is also best for the one-step-ahead. However, as opposed to the total sample investigation, beyond the three-steps-ahead the simple DGSE is better that DGSE-VAR and presents the best performance. The results for the CPI series are different for over two-quarters-ahead compared to those from the total period. Specifically, DGSE-VAR(1) and DSGE-VAR(4) outperform the other models for three- and four-steps-ahead respectively, while BVAR provides with the lowest RMSE for five-quarters-ahead. The results for one- and two-steps are qualitatively the same. Finally, for FFR the TVP-VAR(1) clearly outperforms all other models for all horizons, interestingly ruling out the DSGE-VAR as the best performer in the total out-of-sample period for the four- and five-quarter-ahead. The RMSE ratios for the financial crisis sub-period are reported in Tables 6, 7 and 8.

is taken in billions of real dollars, seasonally adjusted. The CPI is taken as index level, seasonally adjusted. GDP and CPI are taken in first difference logarithmic transformation. The short term interest rate series is constructed as in Clarida et al. (2000). The new sample is taken considering the vintage updates at 2010:4. We intended to investigate whether a longer sample - especially a dataset that goes back that far in the past - offers higher or eventually lower statistical predictability especially for the atheoretical time-varying VAR models, due to structural changes or overfitting and learning of sample idiosyncrasies that do not correspond to contemporary economic conditions.

For the longer sample and for the GDP series the TVP-VAR outperformed the other models. This result is in accordance with the investigation conducted on the original sample. However, as opposed to the original analysis the simple DGSE was in general better than the DSGE-VAR which on average generated the worst forecast ratios, while VARs and BVARs presented similar predictive performance. The DSGE-VAR with any lag structure seemed to be better than VAR and BVARs particularly over the three-quarter ahead forecasts. The results for the CPI series were slightly different compared to that of the initial sample selection. Specifically, the simple DSGE provided the lowest RMSE ratio against the benchmark for all steps-ahead, while the state-space TVP-VAR presented similar predictability as the simple VAR. Overall, the DSGE-VAR and BVAR underperformed relatively to the aforementioned models. Regarding the FFR variable, the TVP-VAR and the VAR models were the winners in case of the long sample as opposed to the TVP-VAR and DSGE-VAR with the original sample. In particular, for the first two quarter-ahead forecasts the TVP-VAR(1) clearly outperformed all other models, while for the three-, four- and five-quarter-ahead the VAR(4) gave the best results against the benchmark. On average the DSGE was better than the DSGE-VAR.

Table 6 RMSE Ratios for the US GDP (2008-2009 sub-period)

GDP	Quarters ahead				
	1	2	3	4	5
VAR(1)	0.970	0.970	0.974	0.967	0.954
VAR(3)	0.990	0.987	0.986	0.983	0.980
VAR(4)	0.977	0.973	0.968	0.962	0.952
BVAR(1)	0.975	0.967	0.981	0.967	0.962
BVAR(2)	1.005	0.999	1.013	1.006	1.015
BVAR(3)	0.994	0.984	0.996	0.990	1.007
BVAR(4)	0.976	0.966	0.981	0.967	0.966
DSGE	0.882	0.860	0.860	0.669	0.782
DSGE-VAR(1)	0.979	0.973	0.983	0.961	0.900
DSGE-VAR(2)	1.012	1.010	0.978	0.879	0.915
DSGE-VAR(3)	1.008	1.018	0.954	0.947	0.944
DSGE-VAR(4)	1.041	1.013	0.927	0.912	0.926
TVP-VAR(1)	0.618	0.822	0.896	0.929	0.951
TVP-VAR(2)	0.706	0.706	0.877	0.879	0.919
TVP-VAR(3)	0.741	0.780	0.824	0.835	0.931
TVP-VAR(4)	0.745	0.819	0.846	0.818	0.812

Notes: The ratios are estimated against the benchmark model VAR(2) for one- to five-steps-ahead in the out-of-sample (rolling) sub-period including the US financial crisis, 2008:1 - 2009:4.

Table 7 RMSE Ratios for the US CPI (2008-2009 sub-period)

CPI	Quarters ahead				
	1	2	3	4	5
VAR(1)	1.004	1.003	1.000	1.002	1.147
VAR(3)	0.993	0.992	0.993	0.991	0.936
VAR(4)	0.990	0.987	0.984	0.985	0.937
BVAR(1)	1.000	1.001	1.000	0.999	1.046
BVAR(2)	1.001	1.004	1.004	1.001	1.041
BVAR(3)	0.994	0.998	1.000	0.994	0.935
BVAR(4)	0.992	0.994	0.991	0.989	0.942
DSGE	0.987	1.012	0.999	0.982	1.475
DSGE-VAR(1)	0.995	1.003	0.973	0.983	1.042
DSGE-VAR(2)	0.997	0.993	0.984	0.982	1.348
DSGE-VAR(3)	0.980	0.977	0.982	0.967	1.547
DSGE-VAR(4)	0.980	0.974	0.976	0.960	1.153
TVP-VAR(1)	1.041	1.002	0.997	0.988	1.394
TVP-VAR(2)	1.138	0.868	0.963	0.974	1.261
TVP-VAR(3)	1.057	0.961	0.979	1.015	1.426
TVP-VAR(4)	1.077	1.040	0.979	1.054	1.590

Notes: As in Table 6

Next, we evaluate the forecast accuracy of the models for the total out-of-sample period by applying a pairwise forecast comparison with the Diebold-Mariano (1995) test based on squared prediction errors and with the Clark and West (2004) test. The results are reported in Tables 9, 10 and 11 in *p-values*. The tests have been conducted on the best performer of each category, namely VAR, BVAR, DSGE-VAR, DSGE and TVP-VAR models and for each examined macro-variable. For example, in case of FFR for the five-quarter-ahead forecasts, the DM and CW tests have been implemented pairwise on

Table 8 RMSE Ratios for the US FFR (2008-2009 sub-period)

CPI	Quarters ahead				
	1	2	3	4	5
VAR(1)	1.052	1.047	1.025	1.050	1.121
VAR(3)	0.907	0.901	0.929	0.908	0.882
VAR(4)	0.869	0.845	0.846	0.852	0.893
BVAR(1)	1.009	1.002	0.974	0.993	1.046
BVAR(2)	1.017	1.003	0.960	0.829	0.908
BVAR(3)	0.924	0.925	0.905	0.849	0.761
BVAR(4)	0.923	0.915	0.898	0.913	0.971
DSGE	0.927	0.973	0.960	0.959	1.084
DSGE-VAR(1)	1.122	1.187	0.859	0.894	0.667
DSGE-VAR(2)	1.014	1.047	0.882	0.712	0.663
DSGE-VAR(3)	0.936	0.944	0.862	0.724	0.666
DSGE-VAR(4)	0.949	0.889	0.763	0.690	0.710
TVP-VAR(1)	0.522	0.431	0.416	0.467	0.493
TVP-VAR(2)	0.562	0.472	0.440	0.486	0.507
TVP-VAR(3)	0.709	0.683	0.638	0.637	0.644
TVP-VAR(4)	0.625	0.600	0.562	0.566	0.571

Notes: As in Table 6

the VAR(3), BVAR(3), DSGE-VAR(1), DGSE and TVP-VAR(4) model specifications. As it was aforementioned the application of the CW test is necessary as the competing models are nested and this creates a bias in the out-of-sample test with the DM test. It uses the Newey-West estimator (1987, 1994) of the asymptotic variance matrix to correct for the autocorrelation of the forecast errors. According to Clark and West (2004) the model with the smaller number of parameters is considered restricted (nested). However, this also depends on the lag specification of the model. In our study we considered the classical VAR as the unrestricted model according to the literature, yet in few cases we account for the total number of parameters and lagged variables when the pairwise comparison includes non-VAR models or vastly different lagged specifications of VARs (e.g., VAR(1) vs. a TVP-VAR (4)). In almost all cases the compared models end-up to contain similar lagged variables leaving only the issue of total number of parameters to lead the consideration of nesting categorization. The application for the CW reveals whether the statistical significance of the DM values vanishes or not. Evidently, as it can be seen below, the results of both tests indicate that in general the forecasts of the investigated models are pairwise significantly different in many cases.

For GDP, the DM test statistics are highly significant at the 1% level for almost all forecasts of the VAR against DSGE-VAR and TVP-VAR as well as for pairs BVAR vs DSGE-VAR and TVP-VAR. The DSGE does not appear to comparatively outperform any of the other models. These compared with the results from the RMSE ratio analysis show that the TVP-VAR and the DSGE-VAR outperform the other models, albeit their pairwise forecast comparison

Table 9 Pairwise forecast comparison for the GDP with Diebold-Mariano and Clark-West tests

GDP	Quarters ahead									
	1		2		3		4		5	
Test	DM	CW	DM	CW	DM	CW	DM	CW	DM	CW
VAR vs BVAR	0.386	0.002	0.083	0.007	0.029	0.002	0.241	0.005	0.696	0.001
VAR vs DSGE-VAR	0.812	0.003	0.033	0.002	0.002	0.163	0.001	0.479	0.013	0.733
VAR vs TVP-VAR	0.008	0.051	0.002	0.057	0.001	0.058	0.002	0.054	0.003	0.048
VAR vs DSGE	0.160	0.123	0.473	0.157	0.140	0.239	0.089	0.268	0.094	0.142
BVAR vs DSGE-VAR	0.505	0.001	0.011	0.009	0.004	0.286	0.006	0.521	0.001	0.134
BVAR vs TVP-VAR	0.007	0.053	0.025	0.069	0.001	0.066	0.005	0.075	0.007	0.052
BVAR vs DSGE	0.169	0.127	0.728	0.158	0.083	0.274	0.275	0.272	0.208	0.145
DSGE-VAR vs TVP-VAR	0.002	0.056	0.004	0.067	0.215	0.073	0.267	0.061	0.859	0.064
DSGE-VAR vs DSGE	0.141	0.123	0.410	0.159	0.811	0.126	0.989	0.203	0.908	0.134
TVP-VAR vs DSGE	0.766	0.014	0.624	0.039	0.800	0.016	0.759	0.036	0.964	0.055

Notes: The results are reported in p-values. The Diebold-Mariano (1995) and Clark-West (2004) tests are based on squared prediction errors. The tests has been conducted on the best performers of each category based on the RMSE results.

Table 10 Pairwise forecast comparison for the CPI with Diebold-Mariano and Clark-West tests

CPI	Quarters ahead									
	1		2		3		4		5	
Test	DM	CW	DM	CW	DM	CW	DM	CW	DM	CW
VAR vs BVAR	0.035	0.002	0.003	0.837	0.004	0.003	0.659	0.002	0.248	0.004
VAR vs DSGE-VAR	0.001	0.566	0.002	0.574	0.001	0.349	0.007	0.009	0.004	0.439
VAR vs TVP-VAR	0.220	0.132	0.741	0.125	0.208	0.120	0.472	0.074	0.003	0.105
VAR vs DSGE	0.025	0.089	0.091	0.082	0.540	0.028	0.047	0.006	0.005	0.109
BVAR vs DSGE-VAR	0.007	0.581	0.004	0.656	0.003	0.545	0.006	0.136	0.003	0.515
BVAR vs TVP-VAR	0.252	0.123	0.153	0.123	0.048	0.119	0.522	0.077	0.001	0.100
BVAR vs DSGE	0.050	0.137	0.751	0.054	0.262	0.005	0.003	0.152	0.009	0.113
DSGE-VAR vs TVP-VAR	0.551	0.130	0.603	0.121	0.865	0.118	0.900	0.115	0.005	0.103
DSGE-VAR vs DSGE	0.392	0.012	0.026	0.117	0.230	0.083	0.254	0.001	0.006	0.140
TVP-VAR vs DSGE	0.965	0.131	0.129	0.128	0.510	0.126	0.517	0.002	0.521	0.116

Notes: As in Table 9

shows no statistical difference especially over the two-quarter-ahead forecasts. Overall, many test statistics are not significant in particular for the BVAR vs DSGE, DSGE-VAR vs TVP-VAR, DSGE-VAR vs DSGE and TVP-VAR vs DSGE pairwise comparisons in all steps-ahead forecasts. After employing the CW test the TVP-VAR pairwise comparisons against VAR and BVAR become statistically weaker, although they retain significance below 10%. With DGSE and DSGE-VAR the differential predictability is stronger, especially over three-steps-ahead. Moreover, the VAR vs DGSE-VAR significance turns weaker while the BVAR vs DGSE-VAR cross-accuracy changes and the VAR vs BVAR forecast comparison is enhanced.

Then, for the CPI series, the five-ahead forecasts seem to be statistically different for all pairs, while VAR and in a smaller extent the BVAR is consistently superior to the other models. In accordance with the RMSE results,

Table 11 Pairwise forecast comparison for the FFR with Diebold-Mariano and Clark-West tests

FFR	Quarters ahead									
	1		2		3		4		5	
Test	DM	CW	DM	CW	DM	CW	DM	CW	DM	CW
VAR vs BVAR	0.002	0.004	0.002	0.003	0.037	0.003	0.002	0.013	0.004	0.820
VAR vs DSGE-VAR	0.005	0.002	0.001	0.001	0.000	0.969	0.003	0.955	0.005	0.831
VAR vs TVP-VAR	0.002	0.001	0.003	0.003	0.002	0.019	0.041	0.103	0.399	0.286
VAR vs DSGE	0.872	0.009	0.002	0.002	0.003	0.006	0.002	0.004	0.004	0.002
BVAR vs DSGE-VAR	0.101	0.060	0.521	0.001	0.003	0.969	0.001	0.946	0.007	0.634
BVAR vs TVP-VAR	0.008	0.000	0.001	0.001	0.001	0.022	0.392	0.091	0.893	0.631
BVAR vs DSGE	0.087	0.385	0.062	0.001	0.005	0.007	0.001	0.003	0.009	0.005
DSGE-VAR vs TVP-VAR	0.004	0.001	0.001	0.001	0.050	0.076	0.942	0.575	0.145	0.825
DSGE-VAR vs DSGE	0.118	0.002	0.432	0.002	0.004	0.002	0.002	0.004	0.002	0.008
TVP-VAR vs DSGE	0.003	0.005	0.002	0.002	0.003	0.004	0.001	0.023	0.001	0.022

Notes: As in Table 9

it is evident that in case of CPI, the VAR setup and the BVAR outperform the other models with statistically significant comparative forecastability at the 1% and 5% level for the DM test. In addition, many statistics appear to be insignificant even at the 10% level. Moreover, the results from CW seem to indicate an increase in statistical predictability of the TVP-VAR against VAR, BVAR and DSGE-VAR. In accordance with the results form DM, the differential predictability of VAR is strengthened. The BVAR and VAR pairs vs DSGE-VAR become weaker, whereas other comparisons appear modified in relation to DM scores. The DGSE vs TVP-VAR comparative performance remains insignificant.

Finally, for the FFR as the combined investigation of the RMSE and DM results indicate, the TVP-VAR in any pair shows a distinctively significant predictability in almost all the steps-ahead forecasts. The DSGE-VAR is also a good performer mainly for the four- and five-ahead forecast. Overall, for the FFR the forecasting ability of the models seems to present a diverse and variant yet significant performance in all steps-ahead. The CW values do not indicate an important change in differential predictability of the TVP-VAR vs VAR, BVAR and DSGE-VAR, although it becomes stronger over the four-quarter-ahead horizon yet not statistically significant in the 10% level. All other results are the same as for the DM test with the exception of the VAR vs DGSE-VAR and BVAR vs DSGE-VAR pairs that turn weaker over three-steps-ahead. Overall, the implementation of the Clark and West (2004) test did not lead to a dramatic weakening in the detected significance by the Diebold-Mariano (1995) test, nor to a distinctive change in most of the pairs. Interestingly, in many cases taking into account the nested relationships lead to a statistically stronger differential predictability[4].

[4] Similarly to the RMSE sensitivity analysis with respect to the sample selection, we performed a robustness check regarding the differential predictability results by the DM and CW tests with the use of a longer sample spanned from 1955:1 to 2009:4 with an out-of-sample period of 2001:1-2009:4. Again, the GDP and the CPI came from the Historical

5 Conclusions

Over the last few years, there has been a growing interest in DSGE modelling in the academia and central banks, in order to explain macroeconomic fluctuations and conduct quantitative policy analysis. Despite their success, DSGE models are not considered the best forecasting tools due to validation, estimation and identification issues. Very recently, hybrid or mixture models have become popular for dealing with some of the DSGE model misspecifications. These models are able to solve the trade-off between theoretical coherence and empirical fit. In this study a DSGE-VAR approach was employed. The main idea behind the DSGE-VAR is the use of the VAR representation as an econometric tool for empirical validation, combining prior information derived from the DSGE model in estimation. However, these models are still linear and they do not consider time-variation for parameters. A novel time-varying multivariate state-space estimation method for vector autoregression models has been introduced in this paper. For the TVP-VAR model, the parameters are estimated using a multivariate specification of the standard Kalman filter (Harvey 1990) combined with a suitable extension of the univariate method-

Data Files for the Real-Time Data Set provided by the Federal Reserve Bank of Philadelphia, while the short interest rate series came from the ALFRED dataset for vintage data provided by the Federal Reserve Bank of St. Louis. As in the initial sample the tests have been conducted on the best performer of each model class and for each macro-variable. For GDP the outcome was almost identical to the original sample selection. The DM test statistics were significant (except for the one-step-ahead) for almost all forecasts of the VAR against DSGE-VAR and TVP-VAR as well as for pairs BVAR vs DSGE-VAR and TVP-VAR. The DM results compared with the RMSE ratio analysis showed that the TVP-VAR and the DSGE-VAR outperformed the other models and their pairwise comparison showed a stronger statistical difference for more steps-ahead than the one of the initial sample. In accordance with the analysis conducted on the original sample, many test statistics were not significant in particular for the BVAR vs DSGE, DSGE-VAR vs TVP-VAR, DSGE-VAR vs DSGE and TVP-VAR vs DSGE pairwise comparisons. After employing the CW test the TVP-VAR pairwise comparisons against the other models became statistically weaker and even more weak than the original sample investigation. Similarly to the initial sample, the VAR vs DGSE-VAR significance turned weaker while the BVAR vs DGSE-VAR and VAR vs BVAR forecast comparison was more significant. For the CPI series in the longer sample, the DM and CW statistics turned weaker compared to the original sample. In this case the VAR setup and the BVAR did not seem to outperform the other models based on the DM test, while the DGSE pairs did not show a statistically significant differential predictability as would someone expect from the RMSE results. In addition, many statistics appeared to be insignificant even at the 10% level, exactly as in the initial sample selection. The results from CW showed the same weak statistical predictability of the TVP-VAR against VAR, BVAR and DSGE-VAR. The DGSE vs TVP-VAR comparative performance remained insignificant. From the combined investigation of the RMSE and DM results for the FFR in the longer sample, it was inferred that the TVP-VAR in any pair showed a distinctively significant predictability in all steps-ahead. The DSGE-VAR was also a good performer for all forecast horizons and not only for the four- and five-ahead forecast as in the original sample. The CW values did not indicate an important change in differential predictability of the TVP-VAR vs VAR, BVAR and DSGE-VAR and in general the statistical significance was stronger compared to the one of the initial sample. Overall, the implementation of the CW test in the longer sample revealed an increase in the statistical significance of the predictability inferred by the DM test concerning the FFR variable, while it did not lead to a dramatic change in most of the pairs for the other two variables.

ology framework of Kim and Nelson (1999). All the aforementioned models as well as standard VARs and Bayesian VARs, are used in a comparative investigation of their out-of-sample predicting performance in case of the GDP, CPI and interest rate series of the US economy. The results vary across the three investigated time series and indicate that, while in general the classical VAR and BVARs provide with equally good forecasting results, in most cases the estimated hybrid DSGE-VAR and the TVP-VAR models outperform the simple DGSE model.

Acknowledgments

The first author acknowledges support from the Marie Curie Intra European Fellowship (FP7-PEOPLE-2009-IEF, N° 251877) under the 7th European Community Framework Programme. The second author acknowledges financial support from "Dote ricercatori": FSE, Regione Lombardia. We are grateful to Helmut Lutkepohl, Massimiliano Marcellino and Tomasz Wozniak for helpful comments. The usual disclaimers apply.

References

1. Adolfson M, Stefan L, Jesper L, Mattias V (2008) Evaluating an Estimated New Keynesian Small Open Economy Model, Journal of Economic Dynamics and Control, 32(8), 2690-2721
2. Adolfson M, Stefan L, Jesper L, Mattias V (2007) Forecasting Performance of an Open Economy DSGE Model, Econometric Reviews, 26(2-4), 289-328
3. Altug S (1989) Time-to-Build and Aggregate Fluctuations: Some New Evidence, International Economic Review, 30(4), 889-920
4. Banbura M, Giannone D, Reichlin L (2010) Large Bayesian vector auto regressions, Journal of Applied Econometrics, 25(1), 71-92
5. Benati L, Surico P (2008) Evolving U.S. Monetary Policy and the decline of inflation predictability, Journal of the European Economic Association 6, 634–646
6. Benati L, Surico P (2009) VAR Analysis and the Great Moderation, American Economic Review, 99(4), 1636-52
7. Boivin J, Giannoni MP (2006) DSGE Models in a Data-Rich Environment, NBER Working Papers 12772
8. Brüggemann R, H. Lütkepohl and M. Marcellino (2008) Forecasting euro area variables with German pre-EMU data, Journal of Forecasting, 27(6), 465–481
9. Canova F, Gambetti L (2004) On the Time Variations of US Monetary Policy: Who is Right?, Money Macro and Finance (MMF) Research Group Conference.
10. Clarida R, Galí J, Gertler M (2000) Monetary Policy Rules And Macroeconomic Stability: Evidence And Some Theory, The Quarterly Journal of Economics, 115(1), 147-180
11. Clark, T.E., M.W. McCracken (2001) Tests of Forecast Accuracy and Encompassing for Nested Models, Journal of Econometrics 105, 85–110.
12. Clark, T.E., K. D. West (2004) Using Out-of-Sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis, Madison, WI: Working Paper, Department of Economics, University of Wisconsin.
13. Christoffel K, Günter C, Anders W (2008) The New Area-Wide Model of the Euro Area - A Micro-Founded Open-Economy Model for Forecasting and Policy Analysis, European Central Bank Working Paper Series n. 944
14. Cogley TF, Sargent TJ (2005) Drift and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S, Review of Economic Dynamics, 8(2), 262-302

15. Cogley T, Primiceri GE, Sargent TJ (2008) Inflation-gap persistence in the U.S., Working Paper 13749, National Bureau of Economic Research 1050 Massachusetts Avenue Cambridge, MA 02138
16. Dahlhaus R (1997) Fitting time series models to nonstationary processes, The Annals of Statistics, 25, 1-37
17. Dahlhaus R (2000) A likelihood approximation for locally stationary processes, The Annals of Statistics, 28, 1762-1794
18. DeJong D, Ingram B, Whiteman C (1996) A Bayesian Approach to Calibration, Journal of Business Economics and Statistics, 14, 1-9
19. De Mol C, Giannone D, Reichlin L (2008) Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?, Journal of Econometrics, 146(2), 318-328
20. Del Negro M, Schorfheide F (2003) Take your model bowling: forecasting with general equilibrium models, Federal Reserve Bank of Atlantic, Economic Review, No. 4th Quarter
21. Del Negro M, Schorfheide F (2004) Priors from General equilibrium Models for VARs, International Economic Review, 45, 643-673
22. Del Negro M, Schorfheide F, Smets F, Wouters R (2007) On the Fit of New-Keynesisan Models, Journal of Business, Economics and Statistics, 25(2), 124-162
23. Diebold FX, Mariano RS (1995) Comparing predictive accuracy, Journal of Business and Economic Statistics, 13(3), 253-263
24. Doan T, Litterman R, Sims C (1984) Forecasting and Conditional Projections Using Realistic Prior Distributions, Econometric Reviews, 3, 1-100
25. Dua P, Subhash CR (1995) A BVAR model for the Connecticut Economy, Journal of Forecasting, 14(3), 167-180
26. Dua P, Smyth DJ (1995) Forecasting U.S. Home Sales Using BVAR Models and Survey Data on Households' Buying Attitudes for Homes, Journal of Forecasting,, 14, 167-180
27. Durbin J, Koopman SJ (2002) Time Series Analysis by State Space Methods, Oxford University Press
28. Edge R, Gurkaynak RS (2011) How Useful are Estimated DSGE Model Forecasts? Finance and Economics Discussion Series 2011-11, Board of Governors of the Federal Reserve System
29. Fernandez-Villaverde J (2009) The Econometrics of DSGE Models, NBER Working Paper 14677
30. Fernández-de-Córdoba G, Torres JL (2010) Forecasting the Spanish Economy with a DSGE Model: An Augmented VAR Approach, Journal of the Spanish Economic Association. Forthcoming
31. Fernandez-Villaverde J, Rubio-Ramirez J, Sargent TJ, Watson MW (2007) A, B, Cs (and Ds) of Understanding VARs, American Economic Review, 97, 3
32. Geweke J (1999) Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication, Econometric Reviews 18(1), 1-126
33. Ghent A (2009) Comparing DSGE-VAR Forecasting Models: How Big are the Differences?, Journal of Economic Dynamics and Control, 2009, 33(4), 864-882
34. Gupta R, Kabundi A (2010) Forecasting Macroeconomic Variables in a Small Open economy: A Comparison between Small- and Large-Scale Models, Journal of Forecasting, 29, 168-185
35. Harvey AC (1990) Forecasting, structural time series and the Kalman filter, Cambridge University Press
36. Harvey AC (1986) Analysis and generalization of a multivariate exponential smoothing model, Management Science, 32, 374-380
37. Herbst E, Schorfheide F (2012) Evaluating DSGE Model Forecasts of Comovements, WP February 24, 2012 (forthcoming in Journal of Econometrics)
38. Ingram B, Whiteman C (1994) Supplanting the Minnesota Prior - Forecasting Macroeconomics Time Series using Real Business Cycle Model Priors, Journal of Monetary Economics, 34, 497-510
39. Ireland P (2004) A Method for Taking Models to the Data, Journal of Economic Dynamics and Control, 28, 1205-1226

40. Justiniano A, Primiceri GE (2008) The Time-Varying Volatility of Macro-economic Fluctuations, American Economic Review, 98(3), 604-41
41. Kadiyala KR, Karlsson S (1997) Numerical Methods for Estimation and Inference in Bayesian VAR-Models, Journal of Applied Econometrics, 12(2), 99-132
42. Kalman RE (1960) A New Approach to Linear Filtering and Prediction. Journal of Basic Engineering, Transactions of the ASME Series D 82, 35-45
43. Kalman RE (1963) New Methods in Wiener Filtering Theory. In: Bogdanoff J.L., Kozin F. (eds.) Proceedings of the First Symposium of Engineering Applications of Random Function Theory and Probability, Wiley, New York, 270-388
44. Kim C-J, Nelson CR (1999) Has The U.S. Economy Become More Stable? A Bayesian Approach Based On A Markov-Switching Model Of The Business Cycle, The Review of Economics and Statistics, MIT Press, 81(4), 608-616
45. Kim C-J, Nelson CR (1999) State space models with Regime switching, MIT Press
46. King RG (2000) The New IS-LM Model: Language, Logic, and Limits, Federal Reserve Bank of Richmond Economic Quarterly, 86, 45-103
47. Kolasa M, Rubaszek M, Skrzypczynski P (2009) Putting the New Keynesian DSGE Model to the Real-Time Forecasting Test, European Central Bank Working Paper Series n. 1110
48. Kryshko M (2010) Data-Rich DSGE and Dynamic Factor Models, Manuscript
49. Kydland FE, Prescott EC (1982) Time to build and aggregate fluctuations, Econometrica, 50(6), 1345-70
50. Litterman RB (1981) A Bayesian Procedure for Forecasting with Vector Autoregressions, Working Paper, Federal Reserve Bank of Minneapolis
51. Litterman RB (1986) Forecasting with Bayesian Vector Autoregressions: Five Years of Experience, Journal of Business and Statistics 4(1), 25–38
52. Liu GD, Gupta R, Schaling E (2009) A New-Keynesian DSGE Model for Forecasting the South African Economy, Journal of Forecasting, 28, 387-404
53. Long JB, Plosser C (1983) Real business cycles, Journal of Political Economics, 91(1), 39-69
54. Lubik TA, Schorfheide F (2004) Testing for Indeterminacy: An Application to U.S. Monetary Policy, American Economic Review, 94(1), 190-217
55. Marcellino, M. (2004) Forecasting EMU macroeconomic variables, International Journal of Forecasting, 20, 359–372
56. McConnell MM, Perez-Quiros G (2000) Output Fluctuations in the United States: What Has Changed since the Early 1980's? American Economic Review, 90(5), 1464-1476
57. Newey, W. K., West, K. D. (1987) A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, Econometrica 55(3), 703–708.
58. Newey, W. K., West, K. D. (1994) Automatic lag selection in covariance matrix estimation, Review of Economic Studies 61(4), 631–654.
59. Prado GH, West M (2001) Bayesian time-varying autoregressions: Theory, methods and applications. Resenhas Journal, 4, 405-422
60. Primiceri GE (2005) Time Varying Structural Vector Autoregressions and Monetary Policy, The Review of Economic Studies, 72, 821-852
61. Ravenna F (2007) Vector Autoregressions and Reduced Form Representations of DSGE models, Journal of Monetary Economics, 54(7), 2048-2064
62. Rubaszek M, Skrzypczynski P (2008) On the Forecasting Perfomance of a Small-Scale DSGE Model, International Journal of Forecasting, 24, 498-512
63. Sargent T (1989) Two Models of Measurements and the Investment Accelerator, Journal of Political Economy, 97(2), 251-287
64. Schorfheide F (2010) Estimation and Evaluation of DSGE Models: Progress and Challenges, University of Pennsylvania
65. Schorfheide F, Sill K, Kryshko M (2010) DSGE model-based forecasting of non-modelled variables, International Journal of Forecasting, 26, 348–373
66. Sims CA (1980) Macroeconomics and reality, Econometrica, 48(1), 1-48
67. Sims CA (1982) Policy analysis with econometric models. Brooking papers on economic activities, 48, 1-48
68. Sims CA (1998) Comment on Glenn Rudebush's 'Do measures of monetary policy in a VAR makes sense?', International Economic Review, 39, 933-948

69. Sims CA (2002) Solving Linear Rational Expectations Models, Computational Economics, 20(1-2), 1-20
70. Sims CA, Zha T (2006) Were There Regime Switches in U.S. Monetary Policy?, American Economic Review, 96(1), 54-81
71. Smets F, Wouters R (2003) An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area, Journal of the European Economic Association, 1, 1123-1175
72. Smets F, Wouters R (2004) Forecasting with a Bayesian DSGE model: an application to the Euro area, Working Paper No. 389, European Central Bank, Frankfurt
73. Spencer DE (1993) Developing a Bayesian Vector Autoregression Forecasting Model, International Journal of Forecasting 9(3), 407–421
74. Stock JH, Watson WM (2001) Vector autoregressions, Journal of Economic Perspectives, 15, 101-115
75. Theil H, Goldberg AS (1961) On Pure and Mixed Estimation in Economics, International Economic Review, 2, 65-78
76. Todd RM (1984) Improving Economic Forecasting with Bayesian Vector Autoregression, Quarterly Review, Federal Reserve Bank of Minneapolis
77. Wang M-C (2009) Comparing the DSGE Model with the Factor Model: An Out-of-Sample Forecasting Experiment, Journal of Forecasting, 28, 167-182
78. Watson MW (1993) Measures of fit for calibrated models, Journal of Political Economy, 101(6), 1011-1040
79. Woodford M (2003) Interest and Prices, Princeton University Press
80. Zellner A (1963) Estimation for seemingly unrelated regression equations: some exact finite sample results, Journal of the American Statistical Association, 58, 977-992