DSGE Model Validation in a Bayesian Framework: an Assessment

Paccagnini, Alessia

Bocconi University

1 May 2010

DSGE Model Validation in a Bayesian Framework: an Assessment

Alessia Paccagnini[1]

May 2010

Abstract

This paper presents the concept of Model Validation applied to a Dynamic Stochastic General equilibrium Model (DSGE). The main problem discussed is the approximation of the statistical representation for a DSGE model when not all endogenous variables are observable. MonteCarlo experiments in artificial world are implemented to assess this problem by using the DSGE-VAR. Two Data Generating Processes are compared: a forward-looking and a backward-looking model. These experiments are followed by an empirical analysis with real world data for the US economy.

JEL CODES: C11, C15, C32

KEYWORDS: Bayesian Analysis, DSGE Models, Vector Autoregressions, MonteCarlo experiments

Non-Technical Summary

Over the last few years, there has been a growing interest in Academia and in Central Banks in using Dynamic Stochastic General Equilibrium Models (DSGE) to explain macroeconomic fluctuations and conduct quantitative policy analysis.

Under an econometric point of view, the performance of a DSGE model is often tested against estimated a Vector Autoregressive model (VAR). This procedure requires that the Data Generating Process (DGP) consistent with the theoretical economic model has a finite-order VAR representation. However, the statistical representation of a DSGE model can be an exact VAR when all the endogenous variables are observable; otherwise a more complex Vector Autoregressive Moving Average (VARMA) is needed. As far as the VARMA representation is concerned, several papers (see Cooley and Dwyer (1998), Chari, Kehoe and McGrattan (2005), Christiano, Eichenbaum and Vigfusson (2006), Ravenna (2007) and Fernandez-Villaverde, Rubio-Ramirez, Sargent and Watson (2007)) have discussed in general terms the conditions to find an infinite-order VAR representation and a finite-order VAR truncation for a VARMA model.

Two main questions are raised in this paper: can we apply the concept of model validation to DSGE models? What happens to model validation when the DSGE model is represented by a VARMA format?

To answer the first question, we introduce the DSGE-VAR (Del Negro and Schorfheide, 2004 and Del Negro, Schorfheide, Smets and Wouters, 2007a) as a new macroeconometric tool for model validation. This mixture model is a hybrid form obtained combining an Unrestricted VAR (UVAR) for actual data and a Bayesian VAR (BVAR) model with dummy observation priors obtained from the statistical representation of the DSGE model. As far as the model validation is concerned, it consists in first giving a statistical representation of the macroeconomic model, then in using this representation to explain the actual data.

The main contribution of this article is the answer to the second question. Two MonteCarlo experiments on artificial data have been designed and performed to check if the DSGE-VAR can be used as a valid instrument for model validation. The first experiment has as DGP the forward-looking model candidate to explain the data in the mixture DSGE-VAR. This model is the same used to generate the restrictions in the BVAR for the theoretical model. The aim of this exercise is to check if the approximation of a VARMA representation can affect the DSGE-VAR use. In the proposed example, the forward-looking model is represented by a VARMA(3,3) which can be restricted by a VAR(2) (see Appendix 2 and Ravenna (2007)). In the second experiment, the DGP is a backward-looking model (Rudebusch and Svensson (1998), Lindé (2001)) with an exact statistical format given by a VAR(3). The aim of this exercise is to check if the hybrid model can recognize the misspecification of the DGP.

The DSGE-VAR model validation procedure is then carried out in the two cases: the expected results are that the null hypothesis of the correct DGP to explain data is not rejected in the first

MonteCarlo experiment, while it is rejected in the second.

To complicate matters, the DSGE-VAR was estimated with VAR lag-lengths from 1 to 8, to allow for misspecification also in the number of lags over the DGP.

The quantitative analysis generates the following findings. First, if the DGP is a forward-looking model, the DSGE-VAR recognizes the DGP, not rejecting the null hypothesis as expected. Second, if the DGP is a backward-looking model, the DSGE-VAR does not recognize the DGP when more lags are added. The null hypothesis is not rejected. Consequently, a DSGE-VAR(∞) tends to be the DSGE model and the data are explain very well, even if there are misspecifications in the DGP. The most important message from these findings is to be careful to approximate a VARMA representation using a finite-order VAR. The experiments are followed by an empirical analysis with the US economy real data. These exercises compare a DSGE-VAR(1) (the most parsimonious representation as suggested by classical Information Criteria) with a DSGE-VAR(4). More lags in the VAR component allow the DSGE-VAR to explain the data better. This problem is given by the fact that the DSGE model can be represented by an infinite VAR.

How can we decide the lag-order? The use of classical Information Criteria can be helpful but they ignore the priors from the DSGE model. Eventually, we can maximize the marginal data density.

The above results contribute to the current discussion on the usefulness of the estimation methods for the DSGE and the application of these models in policy oriented exercises. The DSGE-VAR can be consider a valid instrument for model validation with a careful attention about the statistical representation of the DSGE model.

2

1 Introduction

Over the last few years, there has been a growing interest in Academia and in Central Banks in using Dynamic Stochastic General Equilibrium Models (DSGE) to explain macroeconomic fluctuations and conduct quantitative policy analysis.

Under an econometric point of view, the performance of a DSGE model is often tested against estimated a Vector Autoregressive model (VAR). This procedure requires that the Data Generating Process (DGP) consistent with the theoretical economic model has a finite-order VAR representation. However, the statistical representation of a DSGE model can be an exact VAR when all the endogenous variables are observable; otherwise a more complex Vector Autoregressive Moving Average (VARMA) is needed. As far as the VARMA representation is concerned, several papers (see Cooley and Dwyer (1998), Chari, Kehoe and McGrattan (2005), Christiano, Eichenbaum and Vigfusson (2006), Ravenna (2007) and Fernandez-Villaverde, Rubio-Ramirez, Sargent and Watson (2007)) have discussed in general terms the conditions to find an infinite-order VAR representation and a finite-order VAR truncation for a VARMA model.

Two main questions are raised in this paper: can we apply the concept of model validation to DSGE models? What happens to model validation when the DSGE model is represented by a VARMA format?

To answer the first question, we introduce the DSGE-VAR (Del Negro and Schorfheide, 2004 and Del Negro, Schorfheide, Smets and Wouters, 2007a) as a new macroeconometric tool for model validation. This mixture model is a hybrid form obtained combining an Unrestricted VAR (UVAR) for actual data and a Bayesian VAR (BVAR) model with dummy observation priors obtained from the statistical representation of the DSGE model. As far as the model validation[1] is concerned, it consists in first giving a statistical representation of the macroeconomic model, then in using this representation to explain the actual data.

The main contribution of this article is the answer to the second question. Two MonteCarlo experiments on artificial data have been designed and performed to check if the DSGE-VAR can be used as a valid instrument for model validation. The first experiment has as DGP the forward-looking model candidate to explain the data in the mixture DSGE-VAR. This model is the same used to generate the restrictions in the BVAR for the theoretical model. The aim of this exercise is to check if the approximation of a VARMA representation can affect the DSGE-VAR use. In the proposed example, the forward-looking model is represented by a VARMA(3,3) which can be restricted by a VAR(2) (see Appendix 2 and Ravenna (2007)). In the second experiment, the DGP is a backward-looking model (Rudebusch and Svensson (1998), Lindé (2001)) with an exact statistical format given by a VAR(3). The aim of this exercise is to check if the hybrid model can recognize the misspecification of the DGP.

---

[1]See Consolo, Favero and Paccagnini (2009), Canova and Sala (2009), Komunjer and Ng (2009a, 2009b) and Iskrev (2010) for more details about the issue of the statistical identification of DSGE models.

The DSGE-VAR model validation procedure is then carried out in the two cases: the expected results are that the null hypothesis of the correct DGP to explain data is not rejected in the first MonteCarlo experiment, while it is rejected in the second.

To complicate matters, the DSGE-VAR was estimated with VAR lag-lengths from 1 to 8, to allow for misspecification also in the number of lags over the DGP.

The quantitative analysis generates the following findings. First, if the DGP is a forward-looking model, the DSGE-VAR recognizes the DGP, not rejecting the null hypothesis as expected. Second, if the DGP is a backward-looking model, the DSGE-VAR does not recognize the DGP when more lags are added. The null hypothesis is not rejected. Consequently, a DSGE-VAR(∞) tends to be the DSGE model and the data are explain very well, even if there are misspecifications in the DGP. The most important message from these findings is to be careful to approximate a VARMA representation using a finite-order VAR. The experiments are followed by an empirical analysis with the US economy real data. These exercises compare a DSGE-VAR(1) (the most parsimonious representation as suggested by classical Information Criteria) with a DSGE-VAR(4). More lags in the VAR component allow the DSGE-VAR to explain the data better.

The remainder of the paper is organized as follows. In Section 2, the concept of Model Validation applied to DSGE model is introduced. Section 3, the DSGE-VAR approach proposed by Del Negro and Schorfheide (2004) is discussed as a general assessment and a simple example is presented. In Section 4, results from MonteCarlo experiments in the artificial world are discussed. Section 5 shows the empirical analysis in the real world. Concluding remarks are in Section 6.

2   Model Validation and DSGE

Over the past decade DSGE models have become increasingly popular for their behavioral structure and microfoundations. The DSGE-VAR introduced by Del Negro and Schorfheide (2004) and Del Negro, Schorfheide, Smets and Wouters (2007a) is the main instrument to implement model validation applied to DSGE models. In this mixture model, the DSGE model is treated as a mechanism for generating prior distributions of the parameters. This approach is the final result of a long tradition in macroeconometrics research to estimate and evaluate the economic models.

In the classical macroeconometrics literature (Canova, 1994), two main approaches have been proposed to compare general equilibrium rational expectations models with real data. The first is the estimation approach which is the descendant of the econometric methodology introduced by Haavelmo (1944). The second is the calibration approach introduced by Frisch (1933) and became famous in the macroeconomic literature by Kydland and Prescott (1982).

The two methodologies have several common features (the same strategy in terms of the model specification and solution), but also several differences.

On one side, the estimation attempts to find the parameters of the decision rule that best fit

the data either by the marginal likelihood (ML) or Generalized Method of Moments (GMM). The validity of the specification is implemented by testing restrictions by general goodness of fit tests or by comparing the fit of the two nested models. Besides these two estimation methods, the Vector Autoregression model (VAR) becomes very popular to estimate DSGE models. This technique consists of representing the theoretical model by a pure statistical representation. An example has been introduced by Sims (1980) and is the Unrestricted VAR model, in which a reduced form for the data is used to perform statistical hypothesis. VAR can be taken directly to the data, they are easy to estimate and to generate out-of-sample forecasts.

On the other side, the calibration sets parameters using alternative rules which includes matching long-run averages, using previous microevidence or a priori selection. The most important exercise in the calibrated models is the evaluation of the fit of these models. The concept of model evaluation concerns in selecting a loss function which measures the distance between the set of economic statistics and the set of statistics obtained from the simulated data. Canova (2005) explains that there are essentially four groups of approaches to evaluate DSGE models.

1. Approaches based on R^2-type measure. In Watson (1993), the economic model is viewed as an approximation to the stochastic process which is the data generating process, considering that in the statistical sense the model is not true. The Goodness-of-fit (R^2-type) measure is introduced to provide an assessment evaluation of the approximation. The key ingredient of the measure is the amount of the error needed to be added to the data generated by the model so that the autocovariance implied by the model plus the error match the autocovariance of the observed data.

2. Approaches which measure distance by using the sampling variability of the actual data. For example, the GMM based approach of Christiano and Eichenbaum (1992) or Fève and Langot (1994), the indirect approach of Cecchetti et al. (1993) and the frequency domain approach of Diebold, Ohanian and Berkowitz (1998). This last paper can be considered as an extension of Watson (1993), using a spectral analysis framework.

3. Approaches which measure distance by using the sampling variability of the simulated data, such as the calibration as testing which provides a simple way to judge the distance between population moments or statistics of a simulated macroeconomic model, as in Gregory and Smith (1991). This method has been used by Soderlind (1994) and Cogley and Nason (1994) to evaluate their DSGE models. All these examples take the driving forces as stochastic and the parameters are given. Canova (1994, 1995) and Maffezzoli (2000) allow parameter uncertainty.

4. Approaches which measure distance by using the sampling variability of both actual and simulated data. It is possible to distinguish approaches which allow for variability in the

5

parameters but not in the exogenous processes, such as DeJong et al. (1996, 2000), Geweke (1999) and Schorfheide (2000), or allow both to vary.

Canova (1994) explains that the main difference between the estimation and the calibration is given by the questions the two methodologies ask.

The estimation approach tries to reply to the question: "Given that the model is true, how false is it?", instead the calibration tries to reply to the question: "Given that the model is false, how true is it? ". In the testing model process, an econometrician takes the model seriously as a DGP and analyzes if and what features of the specification are at variance with the data. A calibrationist takes the opposite view: the model, as a DGP for the data, is false. As the sample size grows, the data are generated by the model will have more variance with the observed data. The statistical models rely on the economy theory so loosely, for example, VAR can fail to uncover parameters that are truly structural. This disadvantage may be crucial in policy evaluation exercises, since VAR can exhibit instability across periods when monetary and fiscal policies change. Despite of the popularity in their use, VAR models are subject to multicollinearity problems and can fail to take into account non-linearities in the economy.

The calibrated DSGE models are typically too stylized to be taken directly to the data and often yield fragile results, using traditional econometric methods for estimation (hypothesis testing, forecasting evaluation) (Smets and Wouters, 2003 and Ireland, 2004).

However, these two approaches have several critical points, a third method becomes very popular in the recent literature: the hybrid model. This methodology tries to discipline the concept of model validation applied to DSGE, combining the statistical properties (of the econometrics approach) with information from the economic model (such as the calibration).

The first attempt to combine these two approaches has been introduced by Sargent (1989) and Altug (1989). In these two papers, they propose augmenting a DSGE model with measurement error terms following a first autoregressive process. Ireland (2004) proposed a method similar to Sargent (1989) and Altug (1989), the differences are in imposing no restriction in the measurement errors and the assumption of first-order vector autoregressive for the residuals. Del Negro and Schorfheide (2004) followed this idea of hybrid model proposed by Ireland (2004) presenting the DSGE-VAR. However, the use of a theoretical model to generate a prior for VAR parameters is not new. Theil and Goldberg (1961) introduced the use of dummy observation priors which have been applied to macroeconomics by Litterman (1981), Doan, Litterman and Sims (1984), Ingram and Whiteman (1994), DeJong, Ingram and Whiteman (1996 and 2000), Sims (1996). In these papers, the Bayesian estimation (for VAR or in general for classical regressions) is interpreted as an estimation based on a sample in which the actual data are augmented by an hypothetical sample in which observations are generated by the DSGE model, the so-called dummy observation priors. The basic idea of the use of the dummy observation priors is to express prior beliefs about the parameters. The parameter space can be shrunk by imposing a set of restrictions, which could be

for instance obtained from a theoretical structural model, directly on the parameters. Litterman (1981) and Doan, Litterman and Sims (1984) introduced the so-called Minnesota Prior, where all equations are centered around a random walk with drift. Ingram and Whiteman (1994) was the first contribution which shows that shrinking the VAR estimates towards the restrictions implied by a theoretical model (in this case, a neoclassical Real Business Cycle (RBC) model, see King, Plosser and Rebelo (1988)) produced better forecasting performance than one produced using a Minnesota prior (see Doan, Litterman and Sims (1984)).

In the DSGE-VAR, the theoretical model is used to generate a prior distribution (as theoretical second-order moments) for a structural time-series model that relates the theoretical restrictions of the DSGE. Sims (2007) evidences that one of main advantages of the DSGE-VAR is to define a continuum of models from the theoretical DSGE model to the VAR rather than comparing models at two extremes: the pure statistical VAR and the theoretical DSGE.

In this paper the concept of model validation is applied using the DSGE-VAR, considering its properties to represent an economic model with a statistical representation, taking into account features of the economic model. For its properties, the DSGE-VAR can be used as a valid instrument to implement the model validation concept in the DSGE literature. In the remain Sections, I will illustrate the DSGE-VAR using a simple example in which not all the endogenous variables are observed and I try to understand if the DSGE-VAR can be used as a valid instrument for model validation in this specific case.

3   DSGE-VAR

In Section 3, the DSGE-VAR approach is discussed in a general assessment and by a macroeconomic example.

3.1   DSGE-VAR: A General Assessment

3.1.1   The Likelihood function

The real data are described by the statistical benchmark (proposed by Del Negro and Schorfheide (2004)), an Unrestricted Vector Autoregressive Model (UVAR) of order p:

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

In compact format:

$$Y = X\Phi + U \tag{2}$$

Y is $(T \times n)$ matrix with rows Y_t', X is a $(T \times k)$ matrix $(k = 1 + np, p =$number of lags) with rows $X_t' = [1, Y_{t-1}', ..., Y_{t-p}']$, U is a $(T \times n)$ matrix with rows u_t' and Φ is a $(k \times n) = [\Phi_0, \Phi_1, ..., \Phi_p]'$.

The one step ahead forecast errors u_t have a multivariate normal distribution $N(0, \Sigma_u)$ conditional on past observations of Y.

The log-likelihood function of the data is function of Φ and Σ_u:

$$L(Y|\Phi, \Sigma_u) \propto |\Sigma_u|^{-\frac{T}{2}} x \exp\left\{-\frac{1}{2}tr\left[\Sigma_u^{-1}\left(Y'Y - \Phi'X'Y - Y'X\Phi + \Phi'X'X\Phi\right)\right]\right\} \tag{3}$$

3.1.2   Priors from the Model

The theoretical model, for example a DSGE model, can be represented by using the state-space form solution. Adopting the notation in Fernandez-Villaverde, Rubio-Ramirez, Sargent and Watson (2007):

$$
\begin{aligned}
x_{t+1} &= A(\theta)x_t + B(\theta)\varepsilon_{t+1} \\
y_{t+1} &= C(\theta)x_t + D(\theta)\varepsilon_{t+1}
\end{aligned}
\tag{4}
$$

where ε_t is an $k \times 1$ vector of structural shocks satisfying $E[\varepsilon_t] = 0$, $E[\varepsilon_t \varepsilon_t'] = I$ and $E[\varepsilon_t \varepsilon_{t-j}] = 0$ for $j \neq 0$, x_t is an $n \times 1$ vector of state variables and y_t is a $k \times 1$ vector of variables observed by the econometrician. The matrices A, B, C and D are non-linear functions of the structural parameters in the DSGE model as represented by the vector θ. Estimation and identification can be complicated by the fact that there are more variables than shocks in the system. For simplicity, it is possible to overcome this problem of stochastic singularity, taking the matrix D as a square and invertible matrix, i.e. the number of shocks is equal to the number of observable variables. Komunjer and Ng (2009a , 2009b) provide an identification analysis, generalizing the results shown by Ravenna (2007) to allow the number of shocks to be different from the number of endogenous variables in the model.

In the DSGE-VAR, one of the most important aspects is the finite-order VAR approximation to the DSGE model. Fernandez-Villaverde, Rubio-Ramirez, Sargent and Watson (2007) evidence the necessity to have the eigenvalues of $A - BD^{-1}C$ to be strictly less than one in modulus in order to have y_t with a infinite-order VAR representation given by:

$$y_t = \sum_{j=1}^{\infty} C\left(A - BD^{-1}C\right)^{j-1} BD^{-1}y_{t-j} + D\varepsilon_t \tag{5}$$

However, as argued in Ravenna (2007), the finite order representation will only be exact if all the endogenous state variables are observable and included in the VAR. If the largest eigenvalue is

close to the unity, a VAR with few lags is a poor approximation to the infinite-order VAR implied by the DSGE model.

The finite approximate VAR representation of the solved DSGE model is written as the following VAR in which the number of shocks is equal to the number of the observable variables $\mathbf{Y}_t$:

$$\mathbf{Y}_t = \mathbf{\Phi}_0^*(\theta) + \mathbf{\Phi}_1^*(\theta)\mathbf{Y}_{t-1} + ... + \mathbf{\Phi}_p^*(\theta)\mathbf{Y}_{t-p} + \mathbf{u}_t^* \tag{6}$$

$$\mathbf{u}_t^* \sim N(\mathbf{0}, \Sigma_u^*(\theta))$$

$$\mathbf{Y} = \mathbf{X}\mathbf{\Phi}^*(\theta) + \mathbf{u}^*$$

$$\underset{Txn}{\mathbf{u}^*} = \begin{bmatrix} \underset{Tx1}{\mathbf{u}_1^*} & ... & \underset{Tx1}{\mathbf{u}_n^*} \end{bmatrix} \tag{7}$$

$$\underset{Txn}{\mathbf{Y}} = \begin{bmatrix} \underset{Tx1}{\mathbf{Y}_1} & ... & \underset{Tx1}{\mathbf{Y}_n} \end{bmatrix} \tag{8}$$

$$\underset{Tx(np+1)}{\mathbf{X}} = \begin{bmatrix} \mathbf{X}_1' \\ ... \\ \mathbf{X}_T' \end{bmatrix},$$

$$\underset{1x(np+1)}{\mathbf{X}_t'} = \begin{bmatrix} 1, \underset{1xn}{\mathbf{Y}_{t-1}'} ... \underset{1xn}{\mathbf{Y}_{t-p}'} \end{bmatrix} \tag{9}$$

$$\underset{(np+1)xn}{\mathbf{\Phi}^*(\theta)} = \begin{bmatrix} \underset{nx1}{\mathbf{\Phi}_0^*(\theta)}, \underset{nxn}{\mathbf{\Phi}_1^*(\theta)}, ..., \underset{nxn}{\mathbf{\Phi}_p^*(\theta)} \end{bmatrix}',$$

where all coefficients are convolutions of the structural parameters in the model included in the vector θ. Of course the theoretical model imposes some restrictions on the VAR, that can be tested by evaluating them against the unrestricted VAR. In a series of papers Del Negro and Schorfheide (2004 and 2006) and Del Negro, Schorfheide, Smets and Wouters (2007a) propose a Bayesian framework for model evaluation. This method tilts coefficient estimates of an unrestricted VAR toward the restriction implied by a DSGE model. The weight placed on the DSGE model is controlled by an hyperparameter called λ. This parameter takes values ranging from 0 (no-weight on the DSGE model) to ∞ (no weight on the unrestricted VAR). Therefore, the posterior distribution of λ provides an overall assessment of the validity of the DSGE model restrictions.

The DSGE restrictions are imposed on the VAR by defining:

$$\Gamma_{xx}^*(\theta) = E_\theta^D[\mathbf{X}_t\mathbf{X}_t']$$

$$\Gamma_{xy}^*(\theta) = E_\theta^D[\mathbf{X}_t\mathbf{Y}_t']$$

where E_θ^D defines the expectation with respect to the distribution generated by the DSGE

model, that of course have to be well defined.

Let Γ_{xx}^*, Γ_{yy}^*, Γ_{xy}^* and Γ_{yx}^* be the theoretical second-order moments of the variables in Y and X implied by the DSGE model, where:

$$
\begin{aligned}
\Phi^*\left(\theta\right) &= \Gamma_{xx}^{*-1}\left(\theta\right)\Gamma_{xy}^*\left(\theta\right) \\
\Sigma^*\left(\theta\right) &= \Gamma_{yy}^*\left(\theta\right) - \Gamma_{yx}^*\left(\theta\right)\Gamma_{xx}^{*-1}\left(\theta\right)\Gamma_{xy}^*\left(\theta\right)
\end{aligned}
\tag{10}
$$

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 distribution for the VAR parameters $p(\Phi, \Sigma_u|\theta)$, is of the Inverted-Wishart (IW) and Normal form:

$$
\begin{aligned}
\Sigma_u\,|\theta &\sim IW\left(\lambda T\Sigma_u^*\left(\theta\right), \lambda T - k, n\right) \\
\mathbf{\Phi}\,|\Sigma_u, \theta &\sim N\left(\mathbf{\Phi}^*\left(\theta\right), \frac{1}{\lambda T}\left[\Sigma_u^{-1}\otimes\Gamma_{XX}\left(\theta\right)\right]^{-1}\right)
\end{aligned}
\tag{11}
$$

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 unrestricted VAR and DSGE estimators, large values of λ correspond to small model misspecification and for $\lambda = \infty$ beliefs about DSGE mis-specification degenerate to a point mass at zero. Bayesian estimation could be interpreted as estimation based a sample in which data are augmented by an hypothetical sample in which observations are generated by the DSGE model, the so-called dummy prior observations, within this framework λ determines the length of the hypothetical sample.

3.1.3 Posterior Distribution and Marginal Likelihood Function

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

$$\Sigma_u \,|\, \theta, Y \quad \sim \quad IW\left((\lambda+1)\, T\hat{\Sigma}_{u,b}(\theta), (\lambda+1)\, T - k, n\right) \tag{12}$$

$$\mathbf{\Phi} \,|\, \Sigma_u, \theta, Y \quad \sim \quad N\left(\hat{\mathbf{\Phi}}_b(\theta), \Sigma_u \otimes [\lambda T\Gamma_{XX}(\theta) + \mathbf{X}'\mathbf{X}]^{-1}\right) \tag{13}$$

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

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

where the matrices $\hat{\mathbf{\Phi}}_b(\theta)$ and $\hat{\Sigma}_{u,b}(\theta)$ 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. The equations (12) and (13) show that the smaller λ, the closer the estimates are to the OLS estimates of an unrestricted VAR. Instead, the higher λ 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 $\hat{\Sigma}_{u,b}(\theta)$)

In order to have proper prior distribution (11), λ has to be greater than $\lambda_{\min}$:

$$\lambda_{MIN} \geq \frac{n+k}{T}; k = 1 + p \times n$$

$$p = \text{lags}$$

$$n = \text{endogenous variables}$$

This lambda is considered the minimum lambda. The optimal lambda must be greater or equal than the minimum lambda $\left(\hat{\lambda} \geq \lambda_{MIN}\right)$ to get a non-degenerate prior density, which is a necessary condition for computing meaningful marginal likelihoods.

Adolfson, Laséen, Lindé and Villani (2008) show that λ_{MIN} depends on the model and sample size, hence the marginal likelihood must be reported as a function of the ratio of the number of post-training artificial observations to the number of actual observations, $\hat{\lambda} - \lambda_{MIN}$.

In this paper, the ratio $\frac{\hat{\lambda} - \lambda_{MIN}}{\lambda_{MIN}}$ can be considered as a measure to understand if the DSGE-VAR tends to be well approximated by the DSGE model. If the ratio is high, the distance between the $\hat{\lambda}$ and λ_{MIN} is high, it means the DSGE model can explain the data well. Consequently, this ratio can be conceived as how much the DSGE model explains the actual data over the statistical representation (the VAR) in the hybrid DSGE-VAR.

The posterior simulator used by Del Negro and Schorfheide (2004) is Markov Chain MonteCarlo Method and the used algorithm is the Metropolis-Hastings acceptance method. This procedure generates a Markov Chain from the posterior distribution of θ, Monte Carlo experiments are realized.

See Appendix 1 and Del Negro and Schorfheide (2004) for more details.

The optimal λ is given by maximizing the marginal data density:

$$\widehat{\lambda} = \underset{\lambda \geqslant 0}{\arg\max} L(Y, \lambda)$$

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

3.2   DSGE-VAR: An Example

A simple economic model, a DSGE model with forward-looking features, is chosen as candidate model to explain actual data. This small-scale New- Keynesian represents the actual U.S. time series for real GDP, CPI and Federal Funds Rate over the period 1981:1-2001:4[2]. This forward-looking model is used in one of MonteCarlo experiments in the artificial world.

After the trivariate VAR representation for the real data, the first step consists of defining the likelihood function given by:

$$L(Y|\Phi, \Sigma_u) \tag{14}$$

In the next step, dummy observation priors are generated from the theoretical model.

This economy is made of a representative household with habit persistence. This household maximizes an utility function additive separable in consumption, real money balances and hours worked over infinite lifetime. The household gains utility from consumption relative to the level of technology, real balances of money and disutility from hours worked. The household earns interest from holding government bonds and real profits from the firms. Moreover, the representative household pays lump-sum taxes to the government.

In this economy, a perfectly competitive, representative final goods producer is supposed to use a continuum of intermediate goods as inputs and the prices for these inputs are given. The intermediate good producers are monopolistic firms which uses labour as the only input. The production technology is the same for all the monopolistic firms and fluctuates around the steady-state growth rate. Nominal rigidities are introduced in terms of price adjustment costs for the

---

[2]Del Negro and Schorfheide (2004) consider U.S. quarterly data from 1955:III to 2001:III (1981-2001 is the chosen sample for the estimation).

The data for real output growth come from the Bureau of Economic Analysis (Gross Domestic Product-SAAR, Billions Chained 1996$). The data for inflation come from the Bureau of Labor Statistics (CPI-U: All Items, seasonally adjusted, 1982-1984=100).

GDP and CPI are taken in first difference of logarithmic transformation.

The interest rate series are constructed as in Clarida, Galì and Gertler (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.

monopolistic firms. Each firm maximizes the profits over infinite lifetime by choosing labour input and its price.

The third component in this economy is the government which spends each period a fraction of the total output which fluctuates exogenously. The government issues bonds and levies lump-sum taxes which are the main part in the government's budget constraint.

The last component is the monetary authority which follows the standard Taylor-rule with the inflation target and the output gap.

There are three exogenous economic shocks: the monetary policy shock (in the monetary policy rule), two autoregressive processes, AR(1) which are the government spending and the 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:

$$\widetilde{x}_t = E_t[\widetilde{x_{t+1}}] - \frac{1}{\tau}(\widetilde{R_t} - E_t[\widetilde{\pi_{t+1}}]) + (1 - \rho_g)\widetilde{g_t} + \rho_Z \frac{1}{\tau}\widetilde{z}_t \tag{15}$$

$$\widehat{\pi}_t = \beta E_t[\widetilde{\pi_{t+1}}] + \kappa[\widetilde{x}_t - \widetilde{g}_t] \tag{16}$$

$$\widetilde{R_t} = \rho_R \widetilde{R_{t-1}} + (1 - \rho_R)(\psi_1 \widehat{\pi}_t + \psi_2 \widetilde{x}_t) + \epsilon_{R,t} \tag{17}$$

$$\widetilde{g}_t = \rho_g \widetilde{g_{t-1}} + \epsilon_{g,t} \tag{18}$$

$$\widetilde{z}_t = \rho_z \widetilde{z_{t-1}} + \epsilon_{z,t} \tag{19}$$

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. See details in King (2000) and Woodford (2003).

The rational expectations solution of the linearized model is then computed using the algorithm implemented by Sims (2002) (see Appendix 2, for more details).

The solution of this linearized model is the following policy function that represents the transition equation:

$$\widetilde{\mathbf{Z}}_t = \mathbf{T}(\theta)\widetilde{\mathbf{Z}}_{t-1} + \mathbf{R}(\theta)\varepsilon_t \tag{20}$$

$$\theta = [\kappa, \tau, \psi_1, \psi_2, \rho_R, \rho_g, \rho_z, \sigma_R, \sigma_g, \sigma_Z]' \tag{21}$$

A numerical solution is computed by using the priors reported in Table A1 (see Appendix 2).

$$\mathbf{T}\left(\theta\right) = \begin{bmatrix} 0 & 0 & -3.293 & 0 & 0.7251 & 0.0476 & 0 & 0 \\ 0 & 0 & -0.1551 & 0 & -0.0932 & 0.0169 & 0 & 0 \\ 0 & 0 & 0.3631 & 0 & -0.0246 & 0.0157 & 0 & 0 \\ 0 & 0 & -0.2738 & 0 & -0.0492 & 0.0313 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.8000 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.3000 & 0 & 0 \\ 0 & 0 & -0.1196 & 0 & 0.5882 & 0.0091 & 0 & 0 \\ 0 & 0 & -0.0563 & 0 & -0.0708 & 0.0026 & 0 & 0 \end{bmatrix}$$

$$\mathbf{R}\left(\theta\right) = \begin{bmatrix} -0.6585 & 0.9064 & 0.1587 \\ -0.3102 & -0.1166 & 0.0564 \\ 0.7262 & -0.0308 & 0.0522 \\ -0.5476 & -0.0615 & 0.1045 \\ 0 & 0.63 & 0 \\ 0 & 0 & 0.8750 \\ -0.2391 & 0.7352 & 0.0304 \\ -0.1126 & -0.0885 & 0.0088 \end{bmatrix}$$

Taking into account the system (4), the matrix $A - BD^{-1}C$ is stable since its eigenvalues are strictly less than one in modulus (the largest is 0.9742). See Appendix 2 for more details.

In this example, not all the endogenous variables are observable and the statistical representation of the DSGE model is a VARMA. In Appendix 2, there is an explanation why the statistical representation is a VARMA(3,3). This representation can be restricted using a finite-order VAR:

$$Y^* = X^*\Phi\left(\theta\right) + U^*$$

where Y^*, X^* and U^* derive from the VAR representation for the theoretical model and coefficients matrix $\Phi(\theta)$ is a function of the parameters used in the model. From this representation, dummy observation priors are computed and added to the sample data. Taking into account Bayes Theorem, the posterior is derived, given the prior distribution. After, we maximize the posterior by numerical methods (as described in the general assessment).

To the optimal $\widehat{\lambda}$, a corresponding optimal mixture model is chosen. For example, in Del Negro and Schorfheide (2004), they use the above discussed simple DSGE model as candidate model to explain the real data. US economy time-series for real GDP, CPI and Federal Funds Rate are considered as actual data for 80 observations (from 1981 to 2001). In Del Negro and Schorfheide (2004), the VARMA representation of the DSGE model is reduced by a VAR(4). The found optimal lambda, $\widehat{\lambda}$, is 0.6.

4 MonteCarlo Experiments

The issue concerns the approximation of a VARMA statistical representation by a finite-order VAR is discussed using MonteCarlo experiments. These exercises are implemented in the artificial world, with two different data-generating processes. In the first experiment, the DGP is the forward-looking model used to generate the dummy observation priors, hence the candidate model to explain the data. Since the forward-looking model has not all the endogenous variables observed, the statistical representation is a VARMA. In the second experiment, the DGP is a different macro-economic model, a backward-looking model (Rudebusch and Svensson (1998), Lindé (2001)) which has an exact VAR(3) as statistical representation.

The DSGE-VAR model is implemented in each MonteCarlo exercise to validate the DSGE model.

The MonteCarlo experiments are carried out under the null and alternative hypothesis. Under the null hypothesis, the DGP is the forward-looking model used in the DSGE-VAR to generate priors. Instead, under alternative hypothesis, the DGP is a different model, not the forward-looking model. In the first experiment, we expect not to reject the null hypothesis, since the forward-looking model candidate to generate priors and to explain the data. While, in the second experiment, we expect to reject the null hypothesis since the DGP is a backward-looking model.

How is the DSGE-VAR used to asses these MonteCarlo experiments?

We focus attention on the $\widehat{\lambda}$ and the $\lambda_{\min}$. If the $\widehat{\lambda}$ is greater than the $\lambda_{\min}$ with an high percentage, the forward-looking model explains the data well. The null hypothesis is not rejected. Otherwise, if the $\widehat{\lambda}$ is equal to $\lambda_{\min}$, the forward-looking model is weak to explain the data, the null hypothesis is rejected. The main finding from the artificial world is to be careful to approximate the VARMA representation of the DSGE model, mispecifying the lag-order in the VAR component of the DSGE-VAR. Otherwise, the DSGE can be considered the best representation for the data, even if the DGP is a different model.

4.1 Forward-Looking Data Generating Process

In the first MonteCarlo experiment, the DGP is the forward-looking model explained in Section 3. Since not all the endogenous variables are observable, the statistical representation of the forward-looking model is a VARMA(3,3). Three artificial series[3] are generated and correspond to the three series for Real GDP, CPI and Federal Funds Rate. For each series, 80 observations are generated representing the sample from 1981 to 2001 (the same used in Del Negro and Schorfheide (2004)). The MonteCarlo experiment is carried out and DSGE-VAR approach to compute the optimal λ

---

[3]The artificial data are generated by taking into consideration mean priors for the parameters and for the standard deviations of the shocks reported in Table A1.

is applied. The hybrid model is implemented with a lag-length from 1 to 8. The lambda grid[4] includes the minimum lambda[5] for each VAR representation in the DSGE-VAR. Artificial data are generated by a MonteCarlo experiment with 100 replications[6]. Table 1 summarizes the frequency of the optimal λ for each DSGE-VAR representation (between the brackets (DSGE-VAR()) the number of lags is indicated):

TABLE 1. MonteCarlo experiment with forward-looking data

DSGE-VAR(1)		DSGE-VAR(2)		DSGE-VAR(3)		DSGE-VAR(4)	
$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency
0.09	10	**0.13**	**47**	0.17	23	0.2	7
0.1	**66**	0.17	27	**0.2**	**48**	**0.24**	**37**
0.13	15	0.2	14	0.24	8	0.25	27
0.17	4	0.24	1	0.25	7	0.28	17
0.2	2	0.25	4	0.28	3	0.3	3
0.25	2	0.28	1	0.3	4	0.31	4
0.31	1	0.3	3	0.31	3	0.35	2
		0.31	2	0.35	3	0.4	2
		0.35	1	0.45	1	0.5	1

[4] $\Lambda =$[0 0.09 0.10 0.13 0.17 0.2 0.24 0.25 0.28 0.3 0.31 0.35 0.4 0.45 0.5 0.55 0.6 0.65 0.7 0.8 0.9 1 5 10]. In this set, the bigger lambda is 10. However, in this MonteCarlo experiment, lambda greater than 0.9 are not chosen as optimal lambda.

[5] 0.09 is the minimum lambda in case of VAR(1); 0.13 is the minimum lambda in case of VAR(2); 0.17 is the minimum lambda in case of VAR(3); 0.20 is the minimum lambda in case of VAR(4); 0.24 is the minimum lambda in case of VAR(5); 0.28 is the minimum lambda in case of VAR(6); 0.31 is the minimum lambda in case of VAR(7) and 0.35 is the minimum lambda in case of VAR(8).

[6] In this case, the number of replications in the algorithm to compute the posterior density is 1000. In this exercise, drawing a complete shape for the marginal likelihood function is not the key aim, hence considering less replications could not be a problem. This choice is lead by the fact that realizing a MonteCarlo experiment 100 times needs a great amount of time.

In this experiment, 100 repetitions in the MonteCarlo experiment with 1000 replications in MCMC require around 36 hours with a PC Pentium(R) D CPU 3.00GHz, 0.98 GB for RAM.

TABLE 1. MonteCarlo experiment with forward-looking data

DSGE-VAR(5)		DSGE-VAR(6)		DSGE-VAR(7)		DSGE-VAR(8)	
$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency
0.24	9	0.28	1	0.31	2	0.35	1
0.25	10	0.3	3	**0.35**	**33**	0.4	19
0.28	8	0.31	26	0.4	31	0.45	17
0.3	**28**	**0.35**	**46**	0.45	18	**0.5**	**28**
0.31	16	0.4	18	0.5	4	0.55	23
0.35	21	0.5	1	0.55	2	0.6	3
0.4	2	0.55	2	0.6	3	0.65	6
0.5	1	0.60	1	0.65	3	0.7	2
0.55	1	0.65	1	0.7	1	0.8	1
0.6	1	0.7	1	0.8	1		
0.65	2			0.9	2		
0.7	1						

Table 1 evidences that when lags are added to the VAR component of the DSGE-VAR, the optimal lambda is increasing and becomes bigger than the minimum lambda. The DGP is a VARMA, as shown by Fernandez-Villaverde et al. (2007) and Ravenna (2007), which can be represented by an infinite-order VAR. Since more lags in the VAR, the DSGE-VAR tends to be approximated by a DSGE model. Hence, the candidate model explains the data very well, if we implement a DSGE-VAR(∞). But can we find a restricted finite-order VAR format of the VARMA used as DGP? Table 3 shows the Information Criteria (Akaike (AIC), Schwarz (SIC) and Hannan-Quinn (HQ)). Moreover, Table 2 considers two other relevant informations: what is the optimal lambda across lags for the same MonteCarlo replication and when the ratio $\frac{\widehat{\lambda}-\lambda_{MIN}}{\lambda_{MIN}}$ is decreasing at the first time.

TABLE 2. Summary Table

AIC		SIC		HQ		opt lambda		Ratio	
Lag	Frequency	Lag	Frequency	Lag	Frequency	Lag	Frequency	Lag	Frequency
2	**9**	**2**	**95**	2	41	1	7	**1**	**52**
3	51	3		**3**	**57**	2	18	2	34
4	21			4	2	3	12	3	11
5	11					4	1	4	3
6	2					5	11		
7	2					6	11		
8	2					7	16		
9	1					**8**	**24**		
10	1								

Information Criteria suggest as a possible reduced form for the VARMA representation a finite-order VAR(2). In Appendix 2, following Ravenna (2007), there is a proof that in this example the VARMA(3,3) can be represented by a VAR(2). The optimal lambda is increasing across lags for the same MonteCarlo replication and it is equal to the maximum lag-order in the 24% of the cases. The ratio is decreasing after the first lag in 52% of cases, since at the second lag the difference between the optimal lambda and the minimum lambda is very small. Notice in Table 1, with a DSGE-VAR(2) the optimal lambda is equal to the minimum lambda in 47% of cases. This result depends on the fact that a VAR(2) can be considered as a possible finite-form of the VARMA representation. Moreover in Table 1, when the DSGE-VAR has a VAR with a lag-length greater than 2, the optimal lambda is increasing and we can confirm that adding more lags, the forward-looking model explains the data better.

The same exercise is repeated taking into account the artificial data generated using the DSGE model represented by a VAR(1). In Appendix 3, Tables A2 and Table A3 show that adding more than one lag in the VAR component of the DSGE-VAR, the optimal lambda is increasing and the economic model explains very well the data mispecifying the lag-length.

4.2 Backward-Looking Data Generating Process

In the second MonteCarlo experiment, the null hypothesis is that the DGP is given by the forward-looking model, but under the alternative hypothesis, the DGP is given by a backward-looking model (Rudebusch and Svensson (1998), Lindé (2001)). This backward-looking model is not the candidate model to explain the data, since the dummy observation priors are generated by the forward-looking model. The aim of this exercise is to understand if the DGSE-VAR can discover this misspecification

in the DGP. However, two MonteCarlo exercises are implemented, calibrating the model by using two different parameter sets. In this kind of experiment with a backward-looking model, the DGP is given by an exact VAR with 3 lags, there is not a truncated VARMA representation problem.

The Rudebusch and Svensson model (which is based on the theoretical model presented by Svensson (1997)) has been chosen since it can be considered a good approximation of real data. This model presents a richer dynamic than a simple Svensson model by allowing for four lags of inflation in Phillips Curve (Aggregate Supply, AS) and two lags of output in Aggregate Demand (AD) curve.

The Rudebusch and Svensson model consists of AS and AD equations which explain the output gap (y) (the percentage deviation of output from its steady state level), the inflation rate (π) and the monetary policy (i). A third equation concerns the monetary policy instrument, the short-term interest rate (i) is considered.

The economy is described by the AS and AD equations and an interest rate equation which follows an autoregressive process:

$$\pi_t = \alpha_{\pi_1}\pi_{t-1} + \alpha_{\pi_2}\pi_{t-2} + \alpha_{\pi_3}\pi_{t-3} + \alpha_{\pi_4}\pi_{t-4} + \alpha_y y_{t-1} + \varepsilon_t^\pi \tag{22}$$

$$y_t = \beta_{y_1} y_{t-1} + \beta_{y_2} y_{t-2} + \beta_r \sum_{j=1}^4 \frac{1}{4}(i-\pi)_{t-j} + \varepsilon_t^y \tag{23}$$

$$i_t = \gamma i_{t-1} + \varepsilon_t^i \tag{24}$$

In the AS equation, (22), the annualized inflation rate π depends: on past inflation rates, on the output gap in the previous period and on an exogenous supply shock ε_t^π (i.i.d. with zero mean and constant variance σ_π^2).

In the AD equation, (23), output gap y_t is related to past output gaps y_{t-1} and y_{t-2}, to the average ex post real interest rate in the four previous periods, $\sum_{j=1}^4 \frac{1}{4}(i-\pi)_{t-j}$ and to an exogenous demand shock ε_t^y (i.i.d. with zero mean and constant variance σ_y^2).

The monetary transmission mechanism is via output to inflation rate. In Rudebusch and Svensson model, the sum of the estimated α_{π_j}'s is restricted to be 1 in order to have an acceleration Phillips curve, where long-run monetary neutrality holds.

The interest rate, (24), follows an autoregressive process with an exogenous monetary shock, ε_t^i (i.i.d with zero mean and constant variance σ_i^2).

Rudebusch and Svensson (1998) estimate each equation of the model by using OLS on quarterly US data over the sample period 1961Q1 to 1996Q2.

Lindé (2001) considers the same model, but the parameters estimated for AS-AD come from a MonteCarlo experiment. Lindé estimates the backward-looking model with OLS on the simulated

data from the equilibrium model calibrated with the estimated monetary policy rules.

This last approach is preferred since it is possible to catch significant parameter changes due to the monetary regime shift and Chairman of FED changes (Burns, Volcker, Greenspan).

For generating artificial data, Lindé's estimation for coefficients is used since this calibration provides a stationary VAR representation.

In Appendix 3, Table A4 presents the estimated coefficients proposed by Rudebusch and Svensson and Lindé (see Appendix 3).

In the MonteCarlo experiments of this paper, Lindé's estimation is considered for two samples, the whole sample and the Greenspan sample.

In the first exercise, the estimation of coefficients calibrated for the whole sample is used, generating 80 quarters. In the second exercise, the estimation of coefficients calibrated for the sample in which Greenspan has been Chairman of FED is used. Greenspan sample is around the same period of 80 quarters as considered in Del Negro and Schorfheide (2004) empirical example.

As far as the monetary policy process is concerned, there is no indication about the calibration of the autoregressive coefficient (γ). This coefficient has been estimated on the Federal Funds Rate time series around 0.9 for both the whole sample and the Greenspan sample. Instead, the standard error of the monetary policy shock has been estimated 1.34 for the whole sample and 0.69 for the Greenspan sample.

In this MonteCarlo experiment, we expect to reject the hypothesis that the DGP is given by the candidate forward-looking model. The DSGE-VAR is the instrument used to detect the DGP misspecification.

The backward-looking model has an exact VAR format with 3 lags. The convenient state-space representation is (see Appendix 4, for more details):

$$X_t = AX_{t-1} + v_t \tag{25}$$

The set-up of these experiments is the same of the forward-looking model. We generate 80 observations, the chosen lambda grid[7] considers the minimum lambda for each DSGE-VAR with lag-length from 1 to 8. The artificial data are generated by a MonteCarlo experiment with 100 replications[8].

In the first MonteCarlo, the Whole sample estimation for the parameter is taken into account. Table 3 shows the frequency of $\widehat{\lambda}$ in MonteCarlo experiments.

---

[7]$\Lambda = \{0, 0.09, 0.10, 0.13, 0.17, 0.2, 0.24, 0.25, 0.3, 0.31, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.8, 0.9, 1\}$

[8]In this case, the number of replications in the Metropolis-Hastings algorithm to compute the posterior density is 1000. In this exercise, drawing a complete shape for the marginal likelihood function is not the key aim, hence considering less replications could not be a problem. This choice is lead by the fact that realizing a MonteCarlo experiment 100 times a great amount of time is necessary.

In this experiment, 100 repetitions in the MonteCarlo experiment with 1000 replications in MCMC require around 36 hours with a PC Pentium(R) D CPU 3.00GHz, 0.98 GB for RAM.

TABLE 3. MonteCarlo experiment with backward-looking data (Whole Sample)

DSGE-VAR(1)		DSGE-VAR(2)		DSGE-VAR(3)		DSGE-VAR(4)	
$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency
0.09	**68**	**0.13**	**97**	**0.17**	**71**	**0.2**	**89**
0.1	32	0.17	2	0.2	29	0.24	5
		0.2	1			0.28	6

DSGE-VAR(5)		DSGE-VAR(6)		DSGE-VAR(7)		DSGE-VAR(8)	
$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency
0.24	16	**0.28**	**47**	**0.31**	**84**	**0.35**	**84**
0.25	23	0.3	15	0.35	16	0.4	16
0.28	**41**	0.31	38				
0.3	6						
0.31	14						

In this experiment, we expect to reject the null hypothesis, since the DGP is the backward-looking model. The way to check the null hypothesis is to compare the optimal and the minimum lambda. In Table 3, the minimum lambda is also among the possible $\widehat{\lambda}$ and in any case the percentage of the minimum lambda which is equal to the optimal lambda is very high and impressive. Only in case of DSGE-VAR with 5 lags, the optimal lambda is equal to the minimum lambda with a percentage of 16%. In this exercise, the DSGE-VAR recognizes the misspecification of the DGP.

Table 4 evidences that the classical Information Criteria suggest a VAR(3) as statistical representation for the data. The optimal lambda is increasing across lags and equals to the maximum lag-order in the 94% of the cases. In the majority of the cases, the ratio is decreasing after the fifth lag. These two results about the optimal lambda and the ratio suggest that for small lag-order, the optimal lambda is very near to the minimum lambda

TABLE 4. Summary Table

AIC		SIC		HQ		opt lambda		Ratio	
Lag	Frequency	Lag	Frequency	Lag	Frequency	Lag	Frequency	Lag	Frequency
3	**54**	2	2	**3**	**94**	7	6	1	31
4	11	**3**	**98**	4	5	**8**	**94**	2	3
5	2			5	1			3	18
6	5							4	8
8	1							**5**	**40**
9	10								
10	17								

However, the artificial data have been generated taking estimation for parameters and variance of shocks for a long sample from 1970 to 1997. In this long period, the U.S. economy has faced several monetary policy regimes and crises. The estimation can be affected by these changes in regime, hence an estimation with a small sample can be better. The same procedure is implemented by considering the estimation of coefficients for the period when Greenspan has been the Chairman of FED (1987Q3 to 1997Q2). Moreover, the sample period used to generate the dummy observation priors is from 1981 to 2001, around the Greenspan sample. Results are in Table 5 and 6.

TABLE 5. MonteCarlo experiment with backward-looking data (Greenspan Sample)

DSGE-VAR(1)		DSGE-VAR(2)		DSGE-VAR(3)		DSGE-VAR(4)	
$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency
0.09	**57**	**0.13**	**65**	0.17	33	0.2	24
0.1	43	0.17	29	**0.2**	**53**	0.24	6
		0.2	6	0.24	3	0.25	10
				0.25	3	**0.28**	**40**
				0.28	7	0.3	8
				0.3	1	0.31	12

DSGE-VAR(5)		DSGE-VAR(6)		DSGE-VAR(7)		DSGE-VAR(8)	
$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency
0.24	1	0.28	14	0.31	38	0.35	19
0.25	9	0.3	12	**0.35**	**45**	**0.4**	**77**
0.28	**46**	**0.31**	**69**	0.4	17	0.5	2
0.3	20	0.35	5			0.6	2
0.31	23						
0.35	1						

Table 5 shows that, in every DSGE-VAR exercise, the minimum lambda is also among the possible optimal lambda. In case of DSGE-VAR with only one lag, the possible $\widehat{\lambda}$ is 0.09 or 0.1, since it is always very close to the minimum lambda. The null hypothesis is rejected, the data do not come from the forward-looking model. In case of DSGE-VAR with 2 lags, the optimal lambda is equal to the minimum is 65% cases. The DGP is a VAR(3) and in case of DSGE-VAR with 3 lags, only in 33% cases the minimum lambda is equal to the optimal lambda. Adding lags makes the $\widehat{\lambda}$ bigger than the minimum lambda. Consequently, the DSGE-VAR with more lags tends to be represented by the DSGE model. Misspecifying the lag-order, the DSGE-VAR is not able to recognize the DGP misspecification.

TABLE 6. Summary Table

AIC		SIC		HQ		opt lambda		Ratio	
Lag	Frequency	Lag	Frequency	Lag	Frequency	Lag	Frequency	Lag	Frequency
3	**52**	**3**	**100**	**3**	**96**	2	1	**1**	**31**
4	13			4	2	3	4	2	30
5	3			5	2	4	17	3	12
6	3					5	5	4	27
7	4					7	7		
8	6					**8**	**70**		
9	1								
10	18								

Table 6 shows as a statistical representation for the data a VAR with 3 lags, suggested by the three Information Criteria. The optimal value for lambda is reached at lag 8 in 70% of cases, it means that $\widehat{\lambda}$ is increasing across the lags for the same artificial data. As far as the first fall in the ratio $\frac{\widehat{\lambda}-\lambda_{MIN}}{\lambda_{MIN}}$ is concerned, with a percentage of 31%, the ratio falls after the first lag. However, as pointed out in Table 4, the ratio is zero in the most of cases.

This second MonteCarlo experiment with a backward-looking model evidences that the DSGE-VAR does not help the economist to reject the null hypothesis that DGP is a forward-looking model. In this exercise, there are two sources of misspecification, the DGP and the number of lags in VAR representation. Adding more lags, the optimal lambda is increasing as the power of the economic model to explain the data.

However, the use of the Classical Information Criteria can help the economist to overcome these problems. Since the priors matter, it could be possible to apply a specific information criterion for mixture models to take into account the weight of the priors of the parameter space of the model.

4.3 Comments on Results

The MonteCarlo experiments generate the following findings. First, the DSGE-VAR model tends to be approximate by the DSGE model when adding more lags in the VAR component. In this sense, the DSGE model always explains the data. Second, in case of the forward-looking DGP represented by a VARMA, the DSGE-VAR recognizes that the data-generating process is the same model used for dummy observation priors and a DSGE-VAR(∞) can represent the data better. This result is expected.

Third, in case of the backward-looking DGP, the DSGE-VAR is not always able to recognize that the dummy observation priors are generated by a different model. In this exercise, the DGP

is an exact VAR(3), adding lags more than 3 in the VAR component of the mixture model, the optimal lambda is increasing. This result is not expected and shows how the econometrician should be careful in using the DSGE-VAR. The misspecification of the DGP is not recognized because of the misspecification of the lag-length. Adding more lags in the DSGE-VAR, the theoretical model explains the data better regardless the DGP is the candidate model in the hybrid composition.

In the light of these results, the approximation problem becomes considerable in model validation applied to the DSGE. Suppose to consider a forward-looking model as your candidate model to explain your data. In this model not all the endogenous variables are observable, the statistical representation is a VARMA model that can be truncated by a finite-order VAR (for example a VAR(2)). Another forward-looking model is the DGP. The econometrician implements a DSGE-VAR (5). The outcome is that the economic model explains the data, even if there are several misspecifications.

In this example, a simple New-Keynesian model has been chosen with a VARMA representation which can be reduced using a finite-order VAR(2). With this simple model we get a powerful result: the lag length matters in the DSGE model statistical representation. It is obvious that in case of a more complex model with a VARMA format, a finite-order VAR truncation is needed and the result can become more powerful.

In the DSGE-VAR procedure, additional lags are not penalized in the marginal likelihood which increases with the misspecification of the number of lags. This issue is important for economists which work with the real data and they do not know the DGP. These results depend on the number of replications in the Metropolis-Hastings algorithm used to solve the posterior. It is possible to get a more accurate result, increasing the number of the replications, but this exercise is very time-consuming. Moreover, we can change the variance shocks in the backward-looking model. Actually, when the econometrician takes into consideration a unit variance for shock, the result changes dramatically. The economic model becomes more important in both samples of calibration for the parameters and it seems the artificial data come from the forward-looking model instead of the backward-looking model. See Tables A12-A13 in Appendix 6.

In Appendix 5, forecasting evaluation tables are reported. In these MonteCarlo experiments, the forecasting exercises are useless to choice the best model when a different lag-length is considered.

5   From the MonteCarlo Experiments to the Empirical Analysis

In Section 5, exercises in the US economy real data are implements, in the light of results in the artificial world. As discussed in MonteCarlo experiments, the lambda grid depends on the number of endogenous variables, the number of lags and the number of the observations in the sample size.

In these exercises, we replicate the DSGE-VAR approach with the same forward-looking model presented in Section 3 and the real data used are real GDP, CPI and Federal Funds Rate for the sample 1981-2001. Three exercises are implemented.

The first one is a mere replication of the DSGE-VAR discussed by Del Negro and Schorfheide (2004), with the same features, but the lambda grid. This new lambda set has more values for the hyperparameter lambda, hence it is more precise, $\Lambda = \{0, 0.2, 0.4, 0.5, 0.6, 0.7, 1, 1.4, 1.8, 10\}$[9]. The found optimal lambda is 0.6 (as in Del Negro and Schorfheide, 2004), considering a VAR(4) in the hybrid DSGE-VAR. Table 7 shows the marginal data density with a DSGE-VAR(4).

TABLE 7. DSGE-VAR(4) Sample 1981-2001[10]		
GRID		MDD
0		NaN
0.2		-230.98
0.4		-216.94
0.5		-215.79
0.6		**-215.52**
0.7		-313.72
1		-216.99
1.4		-219.59
1.8		-221.71
10		-335.31
Inf		-242.86

The second exercise consists of increasing the sample size. A large sample from 1961:1 to 2001:4 (160 observations) is taken into account instead of a small sample of 80 observations. The lambda grid changes. The minimum λ is 0.1, instead of 0.2. The new lambda grid is $\Lambda = \{0, 0.10, 0.15, 0.2, 0.4, 0.5, 0.6, 0.7, 1, 1.4, 1.8, 10\}$. In this exercise, the optimal λ is 0.3, instead of 0.6. Table 8 shows the marginal data density with a DSGE-VAR(4) implemented with a large sample.

---

[9]In this case, the infinite lambda representing the case in which restrictions from DSGE model are extremely tight is not considered since in this experiment the analysis of the shape of the marginal likelihood function is not interesting.

[10]25,000 replications are implemented in the Metropolis-Hastings algorithm.

TABLE 8. DSGE-VAR(4) Sample 1961-2001[11]		
GRID		MDD
0		NaN
0.1		-561.36
0.15		-549.46
0.2		-545.72
0.3		**-543.86**
0.4		-544.71
0.5		-545.91
0.6		-547.36
0.7		-549.37
1		-554.10
1.4		-561.16
1.8		-564.20
10		-585.64

We can compare these two results using the ratio $\frac{\widehat{\lambda}-\lambda_{MIN}}{\lambda_{MIN}}$. Whatever the dimension of sample (large or small) is , the ratio is always equal to 2. Consequently, the sample size influences only the minimum and the optimal lambda. The increase of the number of observations does not affect how much the DSGE model can explain the data and how is much the distance between the DSGE-VAR and the two opposite representations: the DSGE model and the pure statistical VAR.

The third exercise consists of changing the lag-length in the VAR component of the hybrid model. MonteCarlo experiments suggest to be careful in approximating a VARMA representation. However, we need to approximate the VARMA by using a finite-order VAR to implement the DSGE-VAR. We can use two alternative methods to assess the lag-length. The first way is to check the lag-order for US real data by calculating classical Information Criteria[12]. Schwarz and Hannan-Quinn criteria suggest the use of only one lag, while Likelihood Ratio, Final Prediction Error and Akaike criteria suggest six lags. The second way is to maximize the marginal data density associated with the DSGE-VAR $(\widehat{\lambda})$, as suggested by Del Negro, Schorfheide, Smets and Wouters (2007b). In this last option, we can evaluate the lag-order, considering into the mixture model, not only the real data, but also the information from the economic model given by priors and the cross-moments. In Table 9, there is the description of the $\widehat{\lambda}$[13] and its marginal data density, considering lags from 1

---

[11]25,000 replications are implemented in the Metropolis-Hastings algorithm.

[12]Likelihood Ratio, Final Prediction Error, Akaike, Schwarz and Hannan-Quinn. It could be interesting Chari, Kehoe, McGrattan (2007) exercise related to Akaike and Schwartz criteria with Structural VAR.

[13]In this case the lambda grid takes into consideration the minimum lambda for each lag, $\Lambda = \{0, 0.09, 0.10, 0.13, 0.17, 0.2, 0.24, 0.25, 0.3, 0.31, 0.35, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$ and 100,000 replications are implemented in the Metropolis-Hastings algorithm.

to 8.

TABLE 9. Maximizing Marginal Data Density

	λ min	λ opt	MDD
DSGE-VAR(1),T=80	**0.09**	**0.13**	**-68.185**
DSGE-VAR(2), T=80	0.13	0.24	-79.161
DSGE-VAR(3), T=80	0.17	0.24	-81.240
DSGE-VAR(4), T=80	0.2	0.24	-89.956
DSGE-VAR(5), T=80	0.24	0.35	-97.151
DSGE-VAR(6), T=80	0.28	0.9	-112.440
DSGE-VAR(7), T=80	0.31	0.6	-106.029
DSGE-VAR(8), T=80	0.35	0.6	-99.786

These results evidence that the marginal data density is maximized with only one lag. Consequently, both using Information Criteria on real data and maximizing the marginal data density, a more parsimonious representation is suggested[14].

Table 10 compares the marginal data density of a DSGE-VAR with a VAR(1) representation for the small and the large sample[15].

---

[14]It is true that considering AIC and SIC on the real data, we forget the total aspect of the DSGE-VAR, supposing that the model is approximated with the same number of the lags of the real data. But the previous results help to use standard model selection criteria as an approximation to choose the number of lags. In this context, it could be very useful to analyze further information criteria, ad hoc in order to recover the prior influence in the DSGE-VAR combination.

[15]In case of small sample with 80 quarters, the new lambda grid is
$\Lambda = \{0, 0.09, 0.10, 0.15, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 1, 1.4, 1.8, 10\}$.
The optimal lambda is 0.15.
In case of large sample with 160 quarters, the new lambda grid is
$\Lambda = \{0, 0.05, 0.09, 0.10, 0.15, 0.2, 0.3, , 0.4, 0.5, 0.6, 0.7, 1, 1.4, 1.8, 10\}$.
The optimal lambda is 0.1.

TABLE 10. VAR(1) Two Samples[16]		
GRID	MDD (T=80)	MDD (T=160)
0	NaN	NaN
0.05		-557.52
0.09	-210.96	-554.70
0.1	-209.61	**-554.06**
0.15	**-208.03**	-554.13
0.2	-208.04	-555.01
0.3	-209.17	-557.36
0.4	-210.51	-559.73
0.5	-212.11	-562.16
0.6	-213.64	-564.22
0.7	-214.85	-565.52
1	-218.42	-570.46
1.4	-223.53	-574.91
1.8	-225.76	-577.25
10	-260.15	-587.35

Table 11 shows a sum-up comparing the DSGE-VAR with one and four lags in the VAR part, for the small and the large sample.

TABLE 11. Summary Table

	λ min	λ opt	λopt-λmin	$(\lambda opt - \lambda min)/\lambda$ min
DSGE-VAR(4), T=80	0.2	0.6	0.4	2
DSGE-VAR(4), T=160	0.1	0.3	0.2	2
DSGE-VAR(1), T=80	0.09	0.15	0.06	0.6667
DSGE-VAR(1), T=160	0.05	0.1	0.05	1

When we implement the DSGE-VAR with one lag, the ratio suggests that the DSGE model explains the data weakly. Adding more lags, the DSGE-VAR can be approximated by the DSGE model. This problem concerns the use of the forward-looking model with not all the endogenous variables observable. The VARMA representation can be approximated by a VAR(∞). The implemented DSGE-VAR is influenced by the VARMA approximation problem. Increasing the lag-order, the DSGE model explains the data better.

In Appendix 5, forecast evaluation tables compare forecasting performance among VAR, DSGE

---

[16]25,000 replications are implemented in Metropolis-Hastings algorithm.

and a DSGE-VAR models. The best performances in terms of forecasting are given by both VAR(1) and VAR(4) with a small sample.

6   Concluding Remarks and Comments

This paper discusses the concept of model validation applied to a DSGE model. When not all the endogenous variables are observable, the statistical representation of a DSGE model is a VARMA format. A bunch of papers (see Cooley and Dwyer (1998), Chari, Kehoe and McGrattan (2005), Christiano, Eichenbaum and Vigfusson (2006), Ravenna (2007) and Fernandez-Villaverde, Rubio-Ramirez, Sargent and Watson (2007)) argued about the conditions to find an infinite-order VAR representation and a finite-order VAR truncation for a VARMA model.

The main instrument used in the model validation is the DSGE-VAR proposed by Del Negro and Schorfheide (2004). In the DSGE-VAR, we need a finite-order VAR.

In this paper two MonteCarlo experiments are implemented. The first exercise has as DGP the forward-looking model candidate to explain the data in the hybrid model, DSGE-VAR. The second exercise has a backward-looking model as DGP.

The first experiment shows how implementing a DSGE-VAR with an increasing lag-length, the DSGE-VAR can be approximated by the DSGE model. Hence, the DSGE explains data better. No surprising result.

The second experiment shows how the DSGE-VAR is not always able to recognize the misspecification in the DGP. Moreover, when a DSGE-VAR with an increasing number of lags is implemented. The main message of this Montecarlo experiment is that mispecifying the lag-length the DSGE-VAR tends to be the DSGE model even if the DGP is a backward-looking model. This result is surprising and not expected.

Artificial world suggests the econometrician to be careful in adding lags in the DSGE-VAR. In the light of these results, exercises in the US economy real data are implemented. This empirical analysis shows how comparing the most parsimonious DSGE-VAR with a DSGE-VAR with four lags, for example, the DSGE model explains the data differently. More explanation of the data if the DSGE-VAR with more lags is implemented. This problem is given by the fact that the DSGE model can be represented by an infinite VAR.

How can we decide the lag-order? The use of classical Information Criteria can be helpful but they ignore the priors from the DSGE model. Eventually, we can maximize the marginal data density.

The above results contribute to the current discussion on the usefulness of the estimation methods for the DSGE and the application of these models in policy oriented exercises. The DSGE-VAR can be consider a valid instrument for model validation with a careful attention about the statistical representation of the DSGE model.

The next steps in the research agenda can be focused analyzing the marginal likelihood function. In this way, we can understand what happens in the DSGE-VAR, adding more lags and we can build a suitable Information Criteria which takes into account the real data and the dummy observation priors.

References

[1] Adolfson Malin, Stefan Laséen, Jesper Lindé and Mattias Villani (2008): "Evaluating an Estimated New Keynesian Small Open Economy Model", *Journal of Economic Dynamics and Control* Elsevier, vol. 32(8), pages 2690-2721.

[2] Altug, Sumru (1989): "Time-to-build and Aggregate Fluctuations: Some New Evidence", *International Economic Review*, 30 (4), pp. 889-920.

[3] Canova, Fabio (1994): "Statistical Inference in Calibrated Models", *Journal of Applied Econometrics*, 9, S123-S144.

[4] Canova, Fabio (1995): "Sensitivity Analysis and Model Evaluation in Simulated Dynamic General Equilibrium Economies", *International Economic Review*, 36, 477-501.

[5] Canova, Fabio (2005): "Methods for Applied Macroeconomic Research ", Princeton University Press, Princeton.

[6] Canova, Fabio and Sala, Luca (2009): "Back to Square One: Identification issues in DSGE models", *Journal of Monetary Economics*, vol. 56(4), pages 431-449.

[7] Cecchetti, Stephen G., Pok-Sang Lam and Nelson C. Mark (1993): "The Equity Premium and the Risk-Free Rate : Matching the Moments", *Journal of Monetary Economics*, vol. 31(1), pages 21-45.

[8] Chari, V.V., Patrick Kehoe and Ellen R. McGrattan (2005): " A Critique of Structural VARs Using Real Business Cycle Theory", Federal Reserve Bank of Minneapolis, Working Paper 631.

[9] Christiano, Lawrence J. and Martin Eichenbaum (1992): "Current Real Business Cycle Theories and Aggregate labor Market Fluctuation", *American Economic Review*, 82, 430-450.

[10] Christiano, Lawrence J., Martin Eichenbaum and Robert Vigfusson (2006): "Assessing Structural VARs", in: Acemoglu, D., Rogoff, K. and Woodford, M. (Eds.), NBER Macroeconomics Annual 2006, Cambridge: The MIT Press.

[11] Clarida, Richard, Jordi Galì and Mark Gertler (2000): "Monetary Policy Rules and Macroeconomics Stability: Evidence and some Theory", *Quarterly Journal of Economics*, 115, pp 147-180.

[12] Consolo, Agostino, Carlo A. Favero and Alessia Paccagnini (2009): "On the Statistical Identification of DSGE Models", *Journal of Econometrics,* 150, 99-115.

[13] Cogley, Timothy and James M. Nason (1994): "Testing the Implications of Long-Run Neutrality for Monetary Business Cycle Models," *Journal of Applied Econometrics*, vol. 9(S), pages S37-70, Suppl. De.

[14] Cooley, Thomas F and Mark Dwyer (1998): "Business Cycle Analysis Without Much Theory. A Look at Structural VARs", *Journal of Econometrics* 83(1-2), 57-88.

[15] Diebold, Francis, Lee Ohanian and Jeremy Berkowitz (1998): "Dynamic Equilibrium Economies: A Framework for Comparing Models and Data", *Review of Economic Studies*, 65, 433-452.

[16] Doan, Thomas, Robert Litterman and Christopher Sims (1984): "Forecasting and Conditional Projections Using Realistic Prior Distributions", *Econometric Reviews*, 3, pp 1-100.

[17] DeJong, David, Beth Ingram and Charles Whiteman (1996):"A Bayesian Approach to Calibration", *Journal of Business Economics and Statistics*, 14, pp 1-9.

[18] DeJong, David, Beth Ingram and Charles Whiteman (2000): "A Bayesian Approach to Dynamic Macroeconomics", *Journal of Econometrics*, 98, pp 203-223.

[19] Del Negro, Marco and Frank Schorfheide (2004): "Priors from General equilibrium Models for VARs", *International Economic Review*, 45, pp 643-673.

[20] Del Negro, Marco and Frank Schorfheide (2006): "How Good Is What You've Got? DSGE-VAR as a Toolkit for Evaluating DSGE Models", *Federal Reserve Bank of Atlanta Economic Review* 91, Second Quarter 2006.

[21] Del Negro, Marco, Frank Schorfheide, Frank Smets and Raf Wouters (2007a): "On the Fit of New-Keynesian Models", *Journal of Business and Economic Statistics*, Volume 25, Number 2 April 2007, pp123-143 .

[22] Del Negro, Marco, Frank Schorfheide, Frank Smets and Raf Wouters (2007b): "On the Fit of New-Keynesian Models: Rejoinder", *Journal of Business and Economic Statistics*, Volume 25, Number 2 April 2007, pp159-162 .

[23] Fève, Patrick and Francois Langot (1994): "The RBC Models Through Statistical Inference: An Application with French Data", *Journal of Applied Econometrics*, 9, S11-S37.

[24] Fernandez-Villaverde, Jesus, Juan Francisco Rubio-Ramirez , Thomas Sargent and Mark Watson (2007): "A, B, C's (and D's) for understanding VARs", *the American Economic Review,* 97, 3, pp 1021–1026.

[25] Frisch, Ragnar (1933): "Propagation Problems and Impulse Problems in Dynamics Economies ", in *Economic Essays in Honor of Gustav Cassel*, London.

[26] Geweke, John (1989): "Bayesian Inference in Econometric Models using MonteCarlo Integration", *Econometrics*, 57, 1317-1339.

[27] Gregory, Allan W. and Gregory Smith (1991): "Calibration as Testing: Inference in Simulated Macro Models", *Journal of Business and Economic Statistics*, 9, 293-303.

[28] Haavelmo, Trigve (1944): "The Probability Approach in Econometrics ", *Econometrica*, 12, Supplement.

[29] Ingram, Beth and Charles Whiteman (1994): "Supplanting the Minnesota Prior - Forecasting Macroeconomics Time Series using Real Business Cycle Model Priors", *Journal of Monetary Economics,* 34,pp 497-510.

[30] Ireland, Peter (2004): "A Method for Taking Models to the Data", *Journal of Economic Dynamics and Control*, 28, pp 1205-1226.

[31] Iskrev, Nikolay (2010): "Local identification in DSGE models", *Journal of Monetary Economics*, vol. 57(2), pages 189-202.

[32] King, Robert G., Charles I. Plosser and Sergio T. Rebelo (1988): "Production, Growth and Business Cycles", *Journal of Monetary Economics*, 21, pp 195-232.

[33] King, Robert G. (2000): "The New IS-LM Model: Language, Logic, and Limits", *Federal Reserve Bank of Richmond Economic Quarterly*, 86, pp 45-103.

[34] Komunjer, Ivana and Serena Ng (2009a): "Issues in the Static Identification of DSGE Models", mimeo.

[35] Komunjer, Ivana and Serena Ng (2009b): "Dynamic Identification of DSGE Models", mimeo.

[36] Kydland, Finn E. and Edward C. Prescott (1982): "Time to build and aggregate fluctuations", *Econometrica*, 50: 1345-13.

[37] Lindé, Jesper (2001): "The Empirical Relevance of Simple Forward- and Backward-Looking Models: A View from a Dynamic General Equilibrium Model", *Sveriges Riksbank Working Paper Series*, December 2001, No 130.

[38] Litterman, Robert B. (1981): "A Bayesian Procedure for Forecasting with Vector Autoregressions", Working Paper, Federal Reserve Bank of Minneapolis.

[39] Maffezzoli, Marco (2000): "Human Capital and International Business Cycle ", *Review of Economic Dynamics*, 3, 137-165.

[40] Ravenna, Federico (2007): "Vector Autoregressions and Reduced Form Representations of DSGE Models", *Journal of Monetary Economics,* 54, pp. 2048-2064.

[41] Rudebusch, Glenn D.R. and Lars E.O. Svensson (1998): "Policy rules for Inflation Targeting,", NBER Working Paper 6512.

[42] Sargent, Thomas (1989): "Two Models of Measurements and the Investment Accelerator", *Journal of Political Economy*, 97 (2), pp. 251-287.

[43] Schorfheide, Frank (2000): "Loss function-based evaluation of DSGE models", *Journal of Applied Econometrics*, 15 (6), pp. 645-670.

[44] Sims, Christopher A. (1980): "Macroeconomics and Reality", *Econometrica*, 48: pp 1-48.

[45] Sims, Christopher A. (1996): "Macroeconomics and Methodology", *Journal of Economic Perspectives*, 10, Winter 1996, 105-120.

[46] Sims, Chistopher. A. (2002): "Solving Linear Rational Expectations Models", *Computational Economics*, 20 (1-2), pp 1-20.

[47] Sims, Christopher A. (2005): "Dummy Observation Priors Revisited", Manuscript, Princeton University.

[48] Sims, Christopher A. (2007): "On the Fit of New-Keynesian Models: Comment", *Journal of Business and Economic Statistics*, Volume 25, Number 2 April 2007.

[49] Smets, Frank and Raf Wouters (2003): "An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area", *Journal of the European Economic Association*, 1, pp 1123-75.

[50] Soderlind, Paul (1994): "Cyclical Properties of a Real Business Cycle Model ", *Journal of Applied Econometrics*, 9, S113-S122.

[51] Svensson, Lars E.O. (1997): "Inflation Forecast Targeting: Implementing and Monitoring Inflation Targets", *European Economic Review*, Vol.41, No. 6, pp. pp 1111-1146.

[52] Theil Henry and Goldberg Arthur S (1961): "On Pure and Mixed Estimation in Economics", *International Economic Review*, 2, pp 65-78.

[53] Watson, Mark (1993): "Measures of Fit for Calibrated Models", *Journal off Political Economy*, 101, 1011-1041.

[54] Woodford, Michael (2003): "*Interest and Prices*", Princeton University Press.

Appendix

Appendix 1: How to generate draws from the posterior distribution of (Φ, Σ_u, θ)

Here we provide the full derivation of the results reported in Section 2 on the DSGE-VAR approach to obtain draws from the posterior distribution of (Φ, Σ_u, θ). The analysis will be conditional to a value for λ which establishes the relevance of the information between the VAR and DSGE in order to estimate the structural parameter θ. We can think of λ as generating a particular model which can support, with a certain degree, the observed data: the marginal data density represents such a measure of goodness and it would help us to discriminate among different models (i.e. different λ).

This appendix describes i) how to compute moments from DSGE models, ii) how to compute a proper prior distribution given such a set of moments conditions, iii) how to derive the marginal data density in case of conjugate prior.

The Bayesian Approach

We follow the Bayesian approach to draw all the relevant inference for the problem at hand. We consider as a good approximation for the vector of observables, $Y_t = (\Delta \ln y_t, \Delta \ln p_t, R_t)'$, an unrestricted Gaussian VAR(p) model for the data.

Together with the likelihood function for the VAR(p) we have to specify a prior distribution for the VAR coefficients. According to Theil and Goldberg (1961) and following the application by Sims (1996), we can recover a prior distribution by using a set of dummy observations. Such a procedure could be seen as a set of restrictions on the VAR(p) coefficients as well. A novelty of the DS approach is to use the DSGE model to derive artificial data, $\left(\tilde{Y}, \tilde{X} \right)$, which can be used to set up the prior.

The VAR model for the data is

$$Y_t = \Phi X_t + E_t, \tag{26}$$

where $X_t = \left[\iota, Y_{t-1}', \ldots, Y_{t-p}' \right]'$ is a vector of dimension $k \times 1$, $k = mp + 1$, which concatenates the constant and p lags of Y_t, and $\Phi = [\Phi_0 \mid \Phi_1 \mid \ldots \Phi_p]$.

The DSGE model can be described by the following state-space representation

$$\tilde{Y}_t = \Lambda_0 (\theta) + \Lambda_1 (\theta) \tilde{Z}_t + V_t, \tag{27}$$

$$\tilde{Z}_t = T (\theta) \tilde{Z}_{t-1} + R (\theta) U_t, \tag{28}$$

which groups the policy function from the RE equilibrium and the mapping between observables,

$\tilde{Y}_t$, and simulated data, $\tilde{Z}_t$. The vector $\tilde{Y}_t$ can be computed by simulation methods with respect to (27) and (28) or analytically since the DSGE model is stationary.

Given the pair of simulated data $\left(\tilde{Y}_t, \tilde{X}_t\right)^{17}$ we can write a similar specification as in (26)

$$\tilde{Y}_t = \Phi \tilde{X}_t + E_t, \tag{29}$$

that indirectly imposes restrictions on Φ driven from the theoretical model; to derive the DSGE-based prior we will construct the likelihood function of the process in (29).

Compute DSGE Moments

Given the state-space representation in (27) and (28), the unconditional variance for $\tilde{Y}_t$ and $\tilde{Z}_t$ are

$$
\begin{aligned}
\Sigma_{z,z} &= T\Sigma_{z,z}T' + R\Sigma_{u,u}R' && (30) \\
\Sigma_{y,y} &= \Lambda_0\Lambda_0' + \Lambda_1\Sigma_{z,z}\Lambda_1' + \Sigma_{v,v} + \Lambda_1 R\Sigma_{u,v} + \Sigma_{u,v}'R'\Lambda_1' && (31)
\end{aligned}
$$

while the unconditional autocorrelation of order k for $\tilde{Y}_t$ reads

$$
\begin{aligned}
\Sigma_{z,z}(k) &= T^k\Sigma_{z,z}(k-1) && (32) \\
\Sigma_{y,y}(k) &= \Lambda_0\Lambda_0' + \Lambda_1\Sigma_{z,z}(k)\Lambda_1' + \Lambda_1\left(T^k\right)R\Sigma_{u,v}. && (33)
\end{aligned}
$$

These high-order second moments matrices will be necessary to construct $\Sigma_{x,x}$ which is a function of the lags of $\tilde{Y}_t$. Here we have omitted the dependence over θ.

Getting a Proper Prior Distribution out of the DSGE model: π_1

The likelihood function for the artificial data reads

$$\mathcal{L}\left(\tilde{Y}; \Phi, \Sigma_e\right) = (2\pi)^{-mT/2} \left|\Sigma_e\right|^{-T/2} \exp\left(-\frac{1}{2}tr\left(\left(\left(\Phi - \tilde{\Phi}\right)\left(\tilde{X}'\tilde{X}\right)\left(\Phi - \tilde{\Phi}\right) + \tilde{S}\right)\Sigma_e^{-1}\right)\right), \tag{34}$$

where the sufficient statistics are,

$$
\begin{aligned}
\tilde{\Phi} &= \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{Y} && (35) \\
\tilde{S} &= \tilde{Y}'\tilde{Y} - \tilde{Y}'\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{Y} && (36)
\end{aligned}
$$

which can be also specified in terms of population moments

$$
\begin{aligned}
\tilde{\Phi} &= \Sigma_{x,x}^{-1}\Sigma_{x,y} && (37) \\
\tilde{S} &= \Sigma_{y,y} - \Sigma_{x,y}'\Sigma_{x,x}^{-1}\Sigma_{x,y} && (38)
\end{aligned}
$$

---

[17] $\tilde{X}_t$ collects lags of $\tilde{Y}_t$.

where, for instance, $\Sigma_{x,y} = E\left(\tilde{X}_t \tilde{Y}_t\right)$.

We thus use a flat prior to construct a proper distribution based on the DSGE model: the Jeffreys prior for the multivariate case reads

$$\pi_0 = |\Sigma_e|^{-\frac{m+1}{2}}. \tag{39}$$

By combining (34) and (39) we get the kernel of the distribution

$$\pi_1 \propto \mathcal{L}\left(\tilde{Y} \mid \Phi, \Sigma_e\right) \times \pi_0, \tag{40}$$

and by integrating with respect to (Φ, Σ_e) we derive the constant of integration

$$P_{\tilde{Y}}\left(\tilde{Y} \mid \theta\right) = (2\pi)^{-m\tilde{v}/2} \times \left|\tilde{S}\right|^{-\frac{\tilde{v}}{2}} \times \left|\tilde{H}\right|^{-\frac{m}{2}} \times \left[2^{m\tilde{v}/2} \times \pi^{m(m-1)/4} \times \Gamma_m\left(\tilde{v}\right)\right], \tag{41}$$

which is needed to have the DSGE-based prior distribution

$$
\begin{aligned}
\pi_1\left(\Phi, \Sigma_e \mid \tilde{Y}, \theta\right) &= \frac{\mathcal{L}\left(\tilde{Y} \mid \Phi, \Sigma_e, \theta\right) \times \pi_0}{P_{\tilde{Y}}\left(\tilde{Y} \mid \theta\right)} \\[2mm]
&= \frac{(2\pi)^{-mT^*/2}}{(2\pi)^{-m\tilde{v}/2}} \times \frac{\left|\tilde{S}\right|^{\tilde{v}/2} \times \left|\tilde{H}\right|^{m/2} \times |\Sigma_e|^{-(T^*+m+1)/2}}{2^{m\tilde{v}/2} \times \pi^{m(m-1)/4} \times \Gamma_m\left(\tilde{v}\right)} \times \\[2mm]
&\quad \exp\left[-\frac{1}{2} tr\left(\tilde{S}\Sigma_e^{-1}\right)\right] \times \exp\left[-\frac{1}{2} tr\left(\left(\Phi - \tilde{\Phi}\right)'\left(\Sigma_{x,x}\right)\left(\Phi - \tilde{\Phi}\right)\Sigma_e^{-1}\right)\right],
\end{aligned} \tag{42}
$$

given $\Sigma_{x,x}$ non-singular and $\tilde{v} \equiv \tilde{T} - k > k + m$.

Hence, $\pi_1\left(\Phi, \Sigma_e \mid \tilde{Y}, \theta\right)$ is distribution from the Normal $\mathcal{N}\left(\tilde{\Phi}, \Sigma_e \otimes H^{-1}\right)$, Inverse-Wishart $\mathcal{IW}\left(\tilde{S}, \tilde{v}\right)$ family.

The Marginal Data Density given: $P\left(Y \mid \theta\right)$

With a proper prior at hand, π_1, we can now combine data and model-based information to fully specify the posterior conditional on the structural parameter θ. By combining the likelihood and the conjugate prior, we get the posterior kernel

$$P_\Phi\left(\Phi, \Sigma_e \mid Y, \theta\right) \propto \mathcal{L}\left(Y \mid \Phi, \Sigma_e\right) \times \pi_1\left(\Phi, \Sigma_e \mid \tilde{Y}, \theta\right), \tag{43}$$

which can be integrated to obtain the marginal data density[18]

$$P_Y\left(Y \mid \theta\right) = (2\pi)^{-Tm/2} \times \frac{\left|\tilde{S}\right|^{\tilde{v}/2}\left|\tilde{H}\right|^{m/2}}{\left|\bar{S}\right|^{\bar{v}/2}\left|\bar{H}\right|^{m/2}} \times \frac{\Gamma_m\left(\bar{v}\right)}{\Gamma_m\left(\tilde{v}\right)} \times 2^{m(\hat{v}+k)/2} \tag{44}$$

The proper posterior reads

$$\begin{aligned}
P_\Phi\left(\Phi, \Sigma_e \mid Y, \theta\right) &= (2\pi)^{-mk/2} \times \left|\Sigma_e\right|^{-k/2} \times \exp\left[-\frac{1}{2}tr\left(\left(\Phi - \bar{\Phi}\right)'\bar{H}\left(\Phi - \bar{\Phi}\right)\Sigma_e^{-1}\right)\right]\dots \\
&\times \frac{\left|\bar{S}\right|^{\bar{v}/2}\left|\bar{H}\right|^{m/2} \times \left|\Sigma_e\right|^{-(\hat{v}+T^*+m+1)/2}}{2^{m\bar{v}/2} \times \pi^{m(m-1)/4}\Gamma_m\left(\bar{v}\right)} \times \exp\left[-\frac{1}{2}tr\left(\bar{S}\Sigma_e^{-1}\right)\right]
\end{aligned} \tag{45}$$

or equivalently

$$\begin{aligned}
p\left(\Phi \mid \Sigma_e; Y, X\right) &= N\left(\bar{\Phi}, \Sigma_e \otimes \bar{H}^{-1}\right) \tag{46} \\
p\left(\Sigma_e \mid Y, X\right) &= IW\left(\bar{S}, \bar{v}\right) \tag{47}
\end{aligned}$$

where the posterior estimates are as follows

- $\bar{H} = X'X + \tilde{T}\Sigma_{x,x}$
- $\bar{\Phi} = \bar{H}^{-1}\left(X'Y + \tilde{T}\Sigma_{x,y}\right)$
- $Q = \hat{\Phi}'\hat{H}\hat{\Phi} + \tilde{\Phi}'\tilde{H}\tilde{\Phi} - \bar{\Phi}'\bar{H}\bar{\Phi}$
- $\bar{S} = \hat{S} + \tilde{S} + Q$
- $\bar{\Sigma}_e = \dfrac{\bar{S}}{\bar{v}}$

Metropolis-Hasting Algorithm

We have obtained the posterior distribution of the VAR coefficients given the structural parameters

$$P\left(\Phi, \Sigma, \theta \mid Y\right) = P_\Phi\left(\Phi, \Sigma \mid Y, \theta\right) \times P_\theta\left(\theta \mid Y\right). \tag{48}$$

We also need to derive the posterior distribution with respect to θ. We use the fact that

$$P_\theta\left(\theta \mid Y\right) \propto \mathbb{K}_\theta\left(\theta \mid Y\right) = P_Y\left(Y \mid \theta\right) \times \pi_2\left(\theta\right) \tag{49}$$

[18]where

$$\begin{aligned}
\tilde{H} &= \left(\tilde{X}'\tilde{X}\right) \\
\bar{v} &= T + \tilde{T} - k
\end{aligned}$$

38

where $P_Y (Y \mid \theta)$ has been computed above and $\pi_2 (\theta)$ is a set of independent prior distributions over each element of the vector of parameters θ; $\mathbb{K}_\theta (\theta \mid Y)$ is the kernel of the posterior. By combining the likelihood and the prior we don't have a closed form solution. We thus need to simulate draws out of the posterior distribution which is unknown. We follow Schorfheide (2000) and Del Negro and Schorfheide (2004) and we implement a Gaussian random walk Metropolis-Hasting algorithm to generate from $P_\theta (\theta \mid Y)$. We set as a scale factor the inverse of the Hessian matrix, $\Sigma_H (\theta)$, with respect to $\mathbb{K}_\theta (\theta \mid Y)$ evaluated at the mode, θ^*. For each candidate draw, $\tilde{\theta}$,

$$\tilde{\theta} = \theta_{s-1} + (\Sigma_H (\theta^*))^{-1/2} N (0, I), \tag{50}$$

we construct an acceptance probability threshold

$$\alpha \left(\tilde{\theta}, \theta_{s-1} \right) = \min \left(1, \frac{\mathbb{K}_\theta \left(\tilde{\theta} \mid Y \right)}{\mathbb{K}_\theta (\theta_{s-1} \mid Y)} \right). \tag{51}$$

If $\alpha \left(\tilde{\theta}, \theta_{s-1} \right)$ is higher than a certain probability (varying for each draw) we accept the draw as coming from the posterior distribution $P_\theta (\theta \mid Y)$ and update the Markov chain $\theta_s = \tilde{\theta}$, otherwise we discard $\tilde{\theta}$ and draw another candidate from (50).

In doing so and by controlling for convergence of the chain, we are able to draw from the posterior distribution of θ. Given the full set of draws, we can thus make inference on any function of the parameters.

Gelfand-Dey Method for $P(Y)$

We compute the marginal data density which consists of integrating out parameters from the posterior distribution to evaluate the set of models: they basically differ from each other from the weight implied by the parameter λ. However, in this case the functional form of the posterior, $P_\theta (\theta \mid Y)$, is not known and therefore we have to rely on simulation methods. To compute $P(Y)$ we use the Gelfand and Dey (1994) method with the correction suggested by Geweke (1999) to avoid problems in the tails of $P(Y)$ which, given the way it is computed, could be not finite.

Once we have a measure of the marginal data density for each model which, in our setup, depends on the choice of λ, we can then compare different models. The idea of comparing different models based on λ clarifies the contribution of the information from the DSGE model in shaping inference. If the maximal of $P(Y)$ is attained for values of λ close to zero, the DSGE model is not strongly supported by the data.

Appendix 2: The Sims (2002) representation of the Forward-Looking model

After linearization around the steady-state, the economy represented by the forward-looking model can be described by the following system of equations:

$$\widetilde{x}_t = E_t[\widetilde{x_{t+1}}] - \frac{1}{\tau}(\widetilde{R}_t - E_t[\widetilde{\pi_{t+1}}]) + (1 - \rho_g)\widetilde{g}_t + \rho_Z\frac{1}{\tau}\widetilde{z}_t \qquad (52)$$

$$\widetilde{\pi}_t = \beta E_t[\widetilde{\pi_{t+1}}] + \kappa[\widetilde{x}_t - \widetilde{g}_t] \qquad (53)$$

$$\widetilde{R}_t = \rho_R\widetilde{R_{t-1}} + (1 - \rho_R)(\psi_1\widetilde{\pi}_t + \psi_2\widetilde{x}_t) + \epsilon_{R,t} \qquad (54)$$

$$\widetilde{g}_t = \rho_g\widetilde{g_{t-1}} + \epsilon_{g,t} \qquad (55)$$

$$\widetilde{z}_t = \rho_z\widetilde{z_{t-1}} + \epsilon_{z,t} \qquad (56)$$

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. See details in King (2000) and Woodford (2003).

The equation (52) is an intertemporal Euler equation obtained from the households' optimal choice of consumption and bond holdings. There is no investment in the model and so output is proportional to consumption and it depends on an exogenous process that can be interpreted as time-varying government spending. The net effects of these exogenous shifts on the Euler equation are captured in the process g_t, which is defined as $\frac{1}{1-\xi_t}$, where ξ_t is the fraction of output consumed by the government. The parameter $\tau > 0$ can be interpreted as the inverse intertemporal elasticity of substitution.

As in Del Negro and Schorfheide (2004), g_t and z_t are assumed to evolve according to univariate AR(1) processes with coefficients ρ_g and ρ_z. The associated iid normal idiosyncratic shocks are $\epsilon_{g,t}$ and $\epsilon_{z,t}$. The standard deviations of these shocks are denoted as σ_g and σ_z.

The equation (53) represents the inflation dynamics determined by the expectional Phillips curve with slope κ. The parameter $0 < \beta < 1$ is the households' discount factor, this parameter could be represented as $\frac{\gamma}{r^*}$, where γ is the steady-state growth rate of technology and $r^*$is the steady-state real interest rate.

The equation (54) describes the behavior of the monetary authority. The central bank follows a nominal interest rate rule by adjusting its instrument to deviations of inflation and output from their respective target levels. The iid normal idiosyncratic shock $\epsilon_{R,t}$ can be interpreted as the unanticipated deviation from the policy rule or as the policy implementation error and ρ_R measures the degree of the central bank's interest rate smoothing. Its standard deviation is denoted by σ_R. The parameters ψ_1 and ψ_2 are the long-run feedback coefficients from the target values of inflation

and output respectively.

In this linearized model, there are three observed endogenous variables (x_t, π_t and R_t), three shocks (ϵ_t^R, ϵ_t^G, ϵ_t^Z), but there are also two unobserved endogenous variables, g_t and z_t. Following Ravenna (2007), if not all the endogenous state variables are observable, it is not possible to find a finite order representation. A VARMA representation is needed when there are unobserved variables.

The solution to the system is the recursive equilibrium law of motion (see Ravenna, 2007 and Fernandez-Villaverde, Rubio-Ramirez, Sargent and Watson, 2007):

$$
\begin{aligned}
y_t &= Cx_{t-1} + Dz_t \\
x_t &= Ax_{t-1} + Bz_t \\
Z(L)z_t &= \varepsilon_t
\end{aligned}
$$

where x_{t-1} is an $n \times 1$ vector of endogenous state variables (in this example is $x_t = [\widetilde{R}_t, \widetilde{g}_t, \widetilde{z}_t]'$,so $n = 3$), z_t is an $m \times 1$ vector of exogenous state variables (monetary, government spending and technology shocks) ($m = 3$), y_t is an $r \times 1$ vector of endogenous variables ($y_t = \left[\widetilde{x}_t, \widetilde{\pi}_t, \widetilde{R}_t\right]'$) ($r = 3$), ε_t is a vector stochastic process of dimension $m \times 1$ ($\varepsilon_t = [\epsilon_{R,t}, \epsilon_{g,t}, \epsilon_{z,t}]'$) ($m = 3$) such that $E(\varepsilon_t) = 0, E(\varepsilon_t \varepsilon_t') = \Sigma, E(\varepsilon_t \varepsilon_s') = 0$ for $s \neq t$ and Σ is a diagonal matrix. $Z(L)$ is the matrix polynomial $[I - Z_1 L - ... - Z_p L^p]$ in the lag operator L defining a stationary vector $AR(p)$ stochastic process. The $Z(L)$ is assumed to be of the first order, $Z(L) = [I - Z_1 L]$.

The first step is to understand if the largest eigenvalue of $A - BD^{-1}C$ is smaller than one in modulus, to check if we can write y_t as infinite-order VAR representation:

$$
y_t = \sum_{j=1}^{\infty} C \left(A - BD^{-1}C\right)^{j-1} BD^{-1} y_{t-j} + D\varepsilon_t
$$

To obtain this result, the log-linearized model is solved using Sims' algorithm (2002) to compute rational expectations.

The rational expectations solution of the linearized model is then computed. The first step towards solution is to cast the model in the following form :

$$
\Gamma_0 \widetilde{\mathbf{Z}}_t = \Gamma_1 \widetilde{\mathbf{Z}}_{t-1} + C + \Psi\epsilon_t + \Pi\eta_t \tag{57}
$$

$t = 1, ..., T$ where C is a vector of constants, ϵ_t is an exogenous vector of shocks, given in this case by $\epsilon_t = [\epsilon_{R,t}, \epsilon_{g,t}, \epsilon_{Z,t}]'$ and η_t is an expectional error, satisfying $E_t(\eta_{t+1}) = 0$, all t. The results are as follows:

$$\widetilde{\mathbf{Z}}_t = \begin{bmatrix} \widetilde{x}_t \\ \widetilde{\pi}_t \\ \widetilde{R}_t \\ \widetilde{R}_t^* \\ \widetilde{g}_t \\ \widetilde{z}_t \\ E_t\widetilde{x_{t+1}} \\ E_t\widetilde{\pi_{t+1}} \end{bmatrix} \qquad \epsilon_t = \begin{bmatrix} \epsilon_t^R \\ \epsilon_t^G \\ \epsilon_t^Z \end{bmatrix} \qquad \eta_t = \begin{bmatrix} \eta_t^x = x_t - E_{t-1}(x_t) \\ \eta_t^\pi = \pi_t - E_{t-1}(\pi_t) \end{bmatrix}$$

$$\Gamma_0 = \begin{bmatrix} 1 & 0 & \frac{1}{\tau} & 0 & -(1-\rho_g) & -\frac{\rho_z}{\tau} & -1 & -\frac{1}{\tau} \\ -\kappa & 1 & 0 & 0 & \kappa & 0 & 0 & -\beta \\ 0 & 0 & 1 & -(1-\rho_R) & 0 & 0 & 0 & 0 \\ -\psi_2 & -\psi_1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\Gamma_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \rho_R & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho_g & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \rho_Z & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\Psi = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \Pi = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Using Table 1 with priors

TABLE A1. Prior Distribution for DSGE Model Parameters for sample 1981-2001

NAME	RANGE	DENSITY	STARTING VALUE	MEAN	SD
$\ln\gamma$	$\mathbb{R}$	$Normal$	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

The solution of this linearized model is the following policy function that represents the transition equation:

$$\widetilde{\mathbf{Z}}_t = \mathbf{T}(\theta)\widetilde{\mathbf{Z}}_{t-1} + \mathbf{R}(\theta)\varepsilon_t \qquad (58)$$

$$\theta = [\kappa,\tau,\psi_1,\psi_2,\rho_R,\rho_g,\rho_z,\sigma_R,\sigma_g,\sigma_Z]'$$

where:

$$\mathbf{T}(\theta) = \begin{bmatrix} 0 & 0 & -3.293 & 0 & 0.7251 & 0.0476 & 0 & 0 \\ 0 & 0 & -0.1551 & 0 & -0.0932 & 0.0169 & 0 & 0 \\ 0 & 0 & 0.3631 & 0 & -0.0246 & 0.0157 & 0 & 0 \\ 0 & 0 & -0.2738 & 0 & -0.0492 & 0.0313 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.8000 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.3000 & 0 & 0 \\ 0 & 0 & -0.1196 & 0 & 0.5882 & 0.0091 & 0 & 0 \\ 0 & 0 & -0.0563 & 0 & -0.0708 & 0.0026 & 0 & 0 \end{bmatrix}$$

$$\mathbf{R}\left(\theta\right) = \begin{bmatrix} -0.6585 & 0.9064 & 0.1587 \\ -0.3102 & -0.1166 & 0.0564 \\ 0.7262 & -0.0308 & 0.0522 \\ -0.5476 & -0.0615 & 0.1045 \\ 0 & 0.63 & 0 \\ 0 & 0 & 0.8750 \\ -0.2391 & 0.7352 & 0.0304 \\ -0.1126 & -0.0885 & 0.0088 \end{bmatrix}$$

The transition equation (58) delivers the dynamics of the deviations of each economic variable from its steady state value. To obtain the dynamics of output, inflation and the policy rate the last equation is combined with the following measurement equation:

$$\mathbf{Z}_t = W(\theta)\widetilde{\mathbf{Z}}_t + D(\theta) + \upsilon_t \tag{59}$$

where

$$\mathbf{Z}_t = \begin{bmatrix} \Delta \ln x_t \\ \Delta \ln P_t \\ \ln R_t^a \end{bmatrix} \tag{60}$$

As in Del Negro and Schorfheide (2004), measurement equations are:

$$\Delta \ln x_t = \ln \gamma + \Delta \widetilde{x}_t + \widetilde{z}_t \tag{61}$$
$$\Delta \ln P_t = \ln \pi^* + \widetilde{\pi}_t$$
$$\ln R_t^a = 4[(\ln r^* + \ln \pi^*) + \widetilde{R}_t]$$

The system is characterized by the following set of parameters:

$$\boldsymbol{\theta} = [\ln \gamma, \ln \pi^*, \ln r^*, \kappa, \tau, \psi_1, \psi_2, \rho_R, \rho_g, \rho_z, \sigma_R, \sigma_g, \sigma_Z]'$$

Taking into account the state, endogenous and observable variables, we can write the matrices:

44

$$A = \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.8 & 0 \\ 0 & 0 & 0.3 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.2510 & 0 & 0 \\ 0 & 0.63 & 0 \\ 0 & 0 & 0.8750 \end{bmatrix}$$

$$C = \begin{bmatrix} -3.293 & 0.7251 & 0.0476 \\ -0.1551 & -0.0932 & 0.0169 \\ 0.3631 & -0.0246 & 0.0157 \end{bmatrix}$$

$$D = \begin{bmatrix} -0.5476 & -0.0615 & 0.1045 \\ 0 & 0.63 & 0 \\ 0 & 0 & 0.8750 \end{bmatrix}$$

$$A - BD^{-1}C = \begin{bmatrix} 0.3222 & 0.3295 & 0.0217 \\ 0.1551 & 0.8932 & -0.0169 \\ -.3631 & 0.0246 & 0.2843 \end{bmatrix}$$

The eigenvalues vector is:

$$eig = \begin{bmatrix} 0.9742 \\ 0.2804 \\ 0.2804 \end{bmatrix}$$

The eigenvalues are all strictly less than one in absolute value, then a infinite-order VAR can be written. The $A - BD^{-1}C$ matrix is said to be stable and the VAR can match up with the theory. The linearized DSGE model is written as:

$$Y_t = FY_{t-1} + Gz_t$$
$$z_t = Z_1 z_{t-1} + \varepsilon_t$$
$$Y_t = \begin{bmatrix} \widetilde{x}_t \\ \widetilde{\pi}_t \\ \widetilde{R}_t \end{bmatrix} ; F = \begin{bmatrix} A & 0 \\ C & 0 \end{bmatrix} ; G = \begin{bmatrix} B \\ D \end{bmatrix}$$

the vector Y_t has dimension $n + r \times 1$ (in this case 6x1). The vector z_t has dimension $m = 3$. Suppose the number of observable variables is equal to the number of shocks ($m = r + n$), if G^{-1}

exists,

$$
\begin{aligned}
Y_t &= FY_{t-1} + Gz_t \\
Gz_t &= (Y_t - FY_{t-1}) \\
z_t &= G^{-1}(Y_t - FY_{t-1}) \text{ then} \\
z_{t-1} &= G^{-1}(Y_{t-1} - FY_{t-2}) \\
&\qquad \text{substituting } z_{t-1} \text{ into} \\
z_t &= Z_1 z_{t-1} + \varepsilon_t \\
z_t &= Z_1(G^{-1}(Y_{t-1} - FY_{t-2})) + \varepsilon_t
\end{aligned}
$$

It is possible to find a restricted VAR(2) representation of the system:

$$
\begin{aligned}
Y_t &= FY_{t-1} + G(Z_1(G^{-1}(Y_{t-1} - FY_{t-2})) + \varepsilon_t) \\
Y_t &= \left(F + GZ_1G^{-1}\right)Y_{t-1} - \left(GZ_1G^{-1}F\right)Y_{t-2} + G\varepsilon_t \\
Y_t &= \Upsilon_1 Y_{t-1} + \Upsilon_2 Y_{t-2} + \eta_t
\end{aligned}
$$

where the VAR innovations $\eta_t = G\varepsilon_t$ are a rotation of the a structural shocks vector ε_t. In this example, $m < r+n$ (3<3+3), the system is singular. To obtain a non-singular VAR representation, some of the observable variables must be dropped from the system to satisfy $m = r+n$. In any cases, a VAR(2) representation (as discussed in Ravenna (2007)) is consistent with the DSGE model.

It is possible to detect the true VARMA representation:

assume $[I - AL]$ is invertible

$$
\begin{aligned}
x_t &= Ax_{t-1} + Bz_t \\
x_t - Ax_{t-1} &= Bz_t \\
x_t[I - AL] &= Bz_t \\
x_{t-1}[I - AL] &= Bz_{t-1} \\
x_{t-1} &= [I - AL]^{-1}BLz_t \\
&\qquad \text{substituting into} \\
y_t &= Cx_{t-1} + Dz_t \\
y_t &= C\left([I - AL]^{-1}BLz_t\right) + Dz_t \\
y_t &= Dz_t + CH(L)BLz_t \tag{62}
\end{aligned}
$$

where $H(L) = [I - AL]^{-1}$ is a lag polynomial of potentially infinite order.

Suppose a VARMA representation for $Z(L) = I$ and $r = m$ (as in this case).

If $z_t = \varepsilon_t$ (62) is a VMA representation of y_t. If D^{-1} is invertible (62) can be written as:

$$y_t = \varsigma_t + CH(L)BD^{-1}L\varsigma_t$$

where $\varsigma_t = D\varepsilon_t = Dz_t$ is the reduced form innovations vector. The inverse matrix of $H(L)$ can be written in terms of its determinant $|H(L)|$ of order n in the lag operator L and the adjoint matrix $V_G(L)$ of order $(n-1)$ in L :

$$H(L)^{-1} = V_G(L)|H(L)|^{-1}$$

Therefore:

$$|H(L)|y_t = |H(L)|\varsigma_t + CV_G(L)BD^{-1}L\varsigma_t = H^*(L)\varsigma_t \tag{63}$$

This equation is a VARMA(n, n). Consequently, the DSGE model has a VARMA(3,3) representation.

Appendix 3: Another Forward-Looking MonteCarlo Experiment

The DGP of the forward-looking artificial data is a VAR(1). Tables A2 and A3 show that adding more lags than the VAR(1), the optimal lambda is increasing and the DSGE-VAR approaches to be the DSGE model with an infinite number of lags.

TABLE A2. MonteCarlo experiment with forward-looking data

DSGE-VAR(1)		DSGE-VAR(2)		DSGE-VAR(3)		DSGE-VAR(4)	
$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency
0.09	27	0.17	28	0.2	2	0.2	2
0.1	**43**	**0.2**	**36**	0.24	12	0.24	1
0.17	19	0.24	11	0.25	19	0.28	8
0.2	7	0.25	14	**0.28**	**20**	0.3	11
0.24	1	0.28	6	0.3	17	0.31	22
0.25	1	0.3	3	0.31	19	**0.35**	**29**
0.31	1	0.4	2	0.35	8	0.4	23
0.35	1			0.4	2	0.6	1
				0.9	1	0.7	1
						0.9	2

DSGE-VAR(5)		DSGE-VAR(6)		DSGE-VAR(7)		DSGE-VAR(8)	
$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency
0.3	4	0.3	1	0.4	2	0.5	8
0.31	2	0.35	3	0.5	28	0.6	10
0.35	19	0.4	21	0.6	20	**0.7**	**48**
0.4	**48**	**0.5**	**45**	**0.7**	**39**	0.8	1
0.5	23	0.6	11	0.8	1	0.9	25
0.6	2	0.7	14	0.9	8	1	8
0.7	1	0.8	1	1	2		
0.9	1	0.9	4				

TABLE A3. Summary Table

AIC		SIC		HQ		opt lambda		Ratio	
Lag	Frequency	Lag	Frequency	Lag	Frequency	Lag	Frequency	Lag	Frequency
1	**93**	**1**	**100**	**1**	**99**	3	1	1	24
2	5			2	1	4	1	**2**	**35**
3	1					6	10	3	26
4	1					7	31	4	15
						8	**57**		

Appendix 4: The Backward-Looking Model (Rudebusch and Svensson, 1998 and Lindé, 2001)

The Rudebusch and Svensson model consists of AS and AD equations which explain the output gap (y) (the percentage deviation of output from its steady state level), the inflation rate (π) and the monetary policy (i). A third equation concerns the monetary policy instrument, the short-term interest rate (i) is considered.

The economy is described by the AS and AD equations and an interest rate equation which follows an autoregressive process:

$$\pi_t = \alpha_{\pi_1}\pi_{t-1} + \alpha_{\pi_2}\pi_{t-2} + \alpha_{\pi_3}\pi_{t-3} + \alpha_{\pi_4}\pi_{t-4} + \alpha_y y_{t-1} + \varepsilon_t^\pi \tag{64}$$

$$y_t = \beta_{y_1}y_{t-1} + \beta_{y_2}y_{t-2} + \beta_r \sum_{j=1}^{4}\frac{1}{4}(i-\pi)_{t-j} + \varepsilon_t^y \tag{65}$$

$$i_t = \gamma i_{t-1} + \varepsilon_t^i \tag{66}$$

The estimation values of the parameter set is presented in Table A4:

TABLE A4. From Rudebusch and Svensson (RS) (1998) and Lindé (2001)

	RS	LINDE'	LINDE'	LINDE'	LINDE'
	Whole sample	Whole sample	Burns	Volcker	Greenspan
	1961Q1-1996Q2	1970Q1-1997Q4	1970Q1-1978Q1	1979Q3-1987Q2	1987Q3-1997Q4
AS					
α_{π_1}	0.7	0.559	0.062	0.136	0.174
α_{π_2}	-0.1	0.293	0.133	0.140	0.077
α_{π_3}	0.28	0.129	0.062	0.051	0.042
α_{π_4}	0.12	0.019	0.041	0.022	0.002
α_y	0.14	0.052	0.496	0.410	-0.003
σ_π		3.46	4.47	5.39	2.65
AD					
β_{y_1}	1.16	0.824	0.474	0.476	0.694
β_{y_2}	-0.25	0.099	0.332	0.327	0.214
β_r	-0.10	-0.015	0.017	-0.041	-0.014
σ_y		2.24	2.83	3.32	2.00

The backward-looking model has an exact VAR with 3 lags. The convenient state-space representation is:

$$X_t = AX_{t-1} + v_t \tag{67}$$

where

$$
X_t =
\begin{bmatrix}
\pi_t \\
y_t \\
i_t \\
\pi_{t-1} \\
\pi_{t-2} \\
\pi_{t-3} \\
y_{t-1} \\
y_{t-2} \\
y_{t-3} \\
i_{t-1} \\
i_{t-2} \\
i_{t-3}
\end{bmatrix}
; X_{t-1} =
\begin{bmatrix}
\pi_{t-1} \\
y_{t-1} \\
i_{t-1} \\
\pi_{t-2} \\
y_{t-2} \\
i_{t-2} \\
\pi_{t-3} \\
y_{t-3} \\
i_{t-3} \\
\pi_{t-4} \\
y_{t-4} \\
i_{t-4}
\end{bmatrix}
; v_t =
\begin{bmatrix}
\varepsilon_t^\pi \\
\varepsilon_t^y \\
\varepsilon_t^i \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{bmatrix}
$$

$$
A =
\begin{bmatrix}
\alpha_{\pi 1} & \alpha_y & 0 & \alpha_{\pi 2} & 0 & 0 & \alpha_{\pi 3} & 0 & 0 & \alpha_{\pi 4} & 0 & 0 \\
-\frac{\beta_r}{4} & \beta_{y1} & -\frac{\beta_r}{4} & -\frac{\beta_r}{4} & \beta_{y2} & -\frac{\beta_r}{4} & -\frac{\beta_r}{4} & 0 & -\frac{\beta_r}{4} & -\frac{\beta_r}{4} & 0 & -\frac{\beta_r}{4} \\
0 & 0 & \gamma & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0
\end{bmatrix}
$$

The three shocks are distributed as a standard normal. In this case, the backward-looking artificial data can be represented by a VAR with 3 lags.

Appendix 5: Forecasting

In this Section, forecasting evaluation is presented for MonteCarlo experiments and for empirical analysis in the real world.

The MonteCarlo analysis is completed by considering the out-of-sample forecasting performance of DSGE-VAR models in artificial world. These models are estimated over the sample from the first quarter of 1981 to the last quarter of 1997 and the out-of-sample performance is used for the period spanning from the first quarter of 1998 to the last quarter of 2001 (16 observations in the forecasting sample). The most used indicator is the Root Mean Squared Error of the forecasting errors from the different models, and is computed as follows:

$$
\begin{aligned}
RMSE^y &= \sqrt{\frac{1}{16} \sum_{h=1}^{16} \left(y_{t+h} - \widehat{y}_{t+h|t}\right)^2} \\
y &= [\Delta \ln x_t, \Delta \ln P_t, \ln R_t]
\end{aligned}
$$

where $\widehat{y}_{t+h|t}$ is the mean forecast computed as the average across draws and $t = 1997 : 4$.

In this case, RMSE for each lag of the different DSGE-VAR model in each replication of the MonteCarlo experiment is computed. In Table A1, for each DSGE-VAR (from 1 to 8 lags), RMSE, the minimum, the maximum and the mean value across the 100 replications in the experiments, for

the three variables (real GDP, CPI, Interest Rate) are reported for forward-looking model ad DGP.

TABLE A5. Forecasting in MonteCarlo experiment with forward-looking as DGP

	MEAN	MAX	MIN		MEAN	MAX	MIN
DSGE-VAR(1)				DSGE-VAR(5)			
$\Delta \ln x_t$	0.66	0.87	0.52	$\Delta \ln x_t$	0.66	0.89	0.48
$\Delta \ln P_t$	0.33	0.42	0.24	$\Delta \ln P_t$	0.32	0.42	**0.21**
$\ln R_t$	**0.94**	1.83	**0.59**	$\ln R_t$	0.99	**1.40**	0.65
DSGE-VAR(2)				DSGE-VAR(6)			
$\Delta \ln x_t$	0.65	0.85	0.51	$\Delta \ln x_t$	0.66	0.85	0.48
$\Delta \ln P_t$	0.34	0.44	0.26	$\Delta \ln P_t$	0.34	0.45	0.24
$\ln R_t$	0.99	1.87	**0.59**	$\ln R_t$	1.36	2.50	0.79
DSGE-VAR(3)				DSGE-VAR(7)			
$\Delta \ln x_t$	**0.64**	0.92	0.46	$\Delta \ln x_t$	0.66	0.85	0.49
$\Delta \ln P_t$	**0.31**	**0.39**	0.23	$\Delta \ln P_t$	0.35	0.46	0.25
$\ln R_t$	0.99	1.51	0.61	$\ln R_t$	1.34	2.27	0.69
DSGE-VAR(4)				DSGE-VAR(8)			
$\Delta \ln x_t$	0.68	0.94	0.47	$\Delta \ln x_t$	0.65	**0.84**	**0.46**
$\Delta \ln P_t$	0.32	0.40	0.25	$\Delta \ln P_t$	0.32	0.43	0.24
$\ln R_t$	1.06	1.73	0.71	$\ln R_t$	1.23	2.01	0.69

Taking into account RMSE for each variable across lags, there is no any clear indication concerning the best forecasting evaluation. However, it seems that a DSGE-VAR with 3 lags has the best forecasting performance for real GDP and CPI and DSGE-VAR with only one lag has the best performance in case of FFR.

In Table A2, the forecasting evaluation in MonteCarlo experiment is presented when parameters of the backward-looking model are calibrated by using the whole sample calibration.

TABLE A6. Forecasting Backward-looking model as DGP (Whole Sample)

	MEAN	MAX	MIN		MEAN	MAX	MIN
DSGE-VAR(1)				DSGE-VAR(5)			
$\Delta \ln x_t$	0.66	0.81	0.51	$\Delta \ln x_t$	0.67	0.84	0.51
$\Delta \ln P_t$	0.33	0.43	0.22	$\Delta \ln P_t$	0.32	0.43	0.24
$\ln R_t$	**0.95**	1.72	0.55	$\ln R_t$	1	**1.66**	0.65
DSGE-VAR(2)				DSGE-VAR(6)			
$\Delta \ln x_t$	0.65	0.86	0.51	$\Delta \ln x_t$	0.67	**0.79**	0.49
$\Delta \ln P_t$	0.34	0.44	0.25	$\Delta \ln P_t$	0.34	0.44	0.21
$\ln R_t$	0.97	1.85	**0.50**	$\ln R_t$	1.37	2.53	0.79
DSGE-VAR(3)				DSGE-VAR(7)			
$\Delta \ln x_t$	0.65	0.87	0.47	$\Delta \ln x_t$	0.65	0.81	0.53
$\Delta \ln P_t$	**0.30**	**0.41**	**0.20**	$\Delta \ln P_t$	0.35	0.46	0.27
$\ln R_t$	1.03	1.83	0.62	$\ln R_t$	1.24	1.98	0.74
DSGE-VAR(4)				DSGE-VAR(8)			
$\Delta \ln x_t$	0.66	0.91	0.50	$\Delta \ln x_t$	**0.64**	0.88	0.41
$\Delta \ln P_t$	0.33	0.44	**0.20**	$\Delta \ln P_t$	0.32	0.42	0.22
$\ln R_t$	1.01	**1.66**	0.60	$\ln R_t$	1.28	2.02	0.59

As before, there is no evidence that a certain model is the best in forecasting performance. In this second table, forecasting evaluation in MonteCarlo experiment is presented where parameters of the backward-looking model are calibrated by using only Greenspan sample calibration.

TABLE A7. Forecasting Backward-looking model as DGP(Greenspan Sample)

	MEAN	MAX	MIN		MEAN	MAX	MIN
DSGE-VAR(1)				DSGE-VAR(5)			
$\Delta \ln x_t$	0.66	0.86	0.53	$\Delta \ln x_t$	0.66	0.84	0.49
$\Delta \ln P_t$	0.33	0.42	0.25	$\Delta \ln P_t$	0.32	0.43	0.23
$\ln R_t$	**0.94**	1.56	0.59	$\ln R_t$	0.97	1.71	0.65
DSGE-VAR(2)				DSGE-VAR(6)			
$\Delta \ln x_t$	0.64	**0.81**	0.48	$\Delta \ln x_t$	0.66	0.90	0.50
$\Delta \ln P_t$	0.34	0.45	0.25	$\Delta \ln P_t$	0.35	0.46	0.27
$\ln R_t$	0.99	2.08	**0.51**	$\ln R_t$	1.36	2.45	0.81
DSGE-VAR(3)				DSGE-VAR(7)			
$\Delta \ln x_t$	0.64	0.92	**0.41**	$\Delta \ln x_t$	0.66	0.83	0.49
$\Delta \ln P_t$	**0.31**	**0.39**	0.23	$\Delta \ln P_t$	0.36	0.45	0.27
$\ln R_t$	0.99	1.50	0.65	$\ln R_t$	1.31	2.51	0.61
DSGE-VAR(4)				DSGE-VAR(8)			
$\Delta \ln x_t$	0.67	0.85	0.44	$\Delta \ln x_t$	**0.63**	0.85	0.47
$\Delta \ln P_t$	0.32	**0.39**	**0.21**	$\Delta \ln P_t$	**0.31**	0.42	0.24
$\ln R_t$	1.05	1.67	0.54	$\ln R_t$	1.15	1.93	0.67

There is no evidence that a certain model is the best in forecasting performance, however, it seems that a DSGE-VAR with 8 lags has the best forecasting performance for real GDP and CPI and DSGE-VAR with only one lag has the best performance in case of FFR.

As done with MonteCarlo experiments, the exercise with real data can be completed by a forecasting evaluation.

The next tables analyze the forecasting performance in the real world. Table A4 shows the forecasting evaluation for the small sample size of 80 quarters from 1981:01 to 2001:04. Instead, Table A5 shows the forecasting evaluation for the large sample of 160 quarters from 1961:01 to 2001:04. The $\widehat{\lambda}$ has been found for each sample is used in this new estimation for the forecasting performance. The sample of the estimation for the forecasting is, respectively, from 1981 to 1997, in case of a small sample (with $\widehat{\lambda}$ estimated for 80 quarters) and from 1961 to 1997, in case of a large sample (with $\widehat{\lambda}$ estimated for 160 quarters).

TABLE A8. The Forecasting Performance of alternative models: small sample				
MODEL		$\Delta \ln x_t$	$\Delta \ln P_t$	$\ln R_t$
		$RMSE$	$RMSE$	$RMSE$
VAR(4) 80Q		0.62	0.29	0.87
DSGE		0.62 (1.00)	0.27 (0.93)	**0.72** (0.83)
DSGE-VAR(4)($\lambda^* = 0.6$) 80Q		**0.61** (0.98)	**0.26** (0.90)	0.80 (0.92)

RMSE relative to the VAR(4) within brackets

TABLE A9. The Forecasting Performance of alternative models: large sample				
MODEL		$\Delta \ln x_t$	$\Delta \ln P_t$	$\ln R_t$
		$RMSE$	$RMSE$	$RMSE$
VAR(4) 160Q		0.65	0.26	0.78
DSGE		**0.62** (0.95)	0.27 (1.04)	**0.72** (0.92)
DSGE-VAR(4)($\lambda^* = 0.3$) 160Q		0.83 (1.28)	**0.26** (1.00)	0.77 (0.99)

RMSE relative to the VAR(4) within brackets

Considering the results shown in Tables A4-A5, the hybrid model, DSGE-VAR does not seem to be always the best model in terms of the forecasting performance. Moreover, it is interesting to note that using a larger sample in DSGE-VAR model does not improve as one would expect the forecast performance. This aspect may depend on the priors that have been considered to be the same for all the samples. However, in case of DSGE, the interest rate has the best forecast performance in both samples.

In Table A6-A7, the forecasting performance is evaluated by considering the most parsimonious model with only one lag for VAR and DSGE-VAR. Moreover, DSGE-VAR are evaluated on the two different samples: the small sample and the large one.

TABLE A10. The Forecasting Performance of alternative models: VAR(1), small sample			
MODEL	$\Delta \ln x_t$	$\Delta \ln P_t$	$\ln R_t$
	$RMSE$	$RMSE$	$RMSE$
VAR(1)	0.62	0.29	0.90
DSGE	0.62 (1.00)	0.27 (0.93)	**0.72** (0.80)
DSGE-VAR(1)($\lambda^* = 0.15$) 80Q	**0.59** (0.95)	**0.26** (0.90)	1.04 (1.16)

RMSE relative to the VAR(1) within brackets

TABLE A11. The Forecasting Performance of alternative models: VAR(1), large sample			
MODEL	$\Delta \ln x_t$	$\Delta \ln P_t$	$\ln R_t$
	$RMSE$	$RMSE$	$RMSE$
VAR(1)	**0.62**	0.38	0.97
DSGE	**0.62** (1.00)	0.27 (0.71)	**0.72** (0.74)
DSGE-VAR(1)($\lambda^* = 0.1$) 160Q	0.63 (1.02)	**0.26** (0.68)	1.41 (1.45)

RMSE relative to the VAR(1) within brackets

In this case, the best performed model is the DSGE-VAR (apart for FFR) representation, in case of a small sample; instead in case of a larger sample, there is no models with the best forecast performance. However, as before, RMSE for the interest rate is the smallest one in case of DSGE model.

Appendix 6: Other Results

TABLE A12. MonteCarlo experiment with backward-looking data (Whole Sample) unitary variance shock

DSGE-VAR(1)		DSGE-VAR(2)		DSGE-VAR(3)		DSGE-VAR(4)	
$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency
0.09	**76**	**0.2**	**40**	0.2	1	0.25	2
0.1	8	0.24	11	0.24	7	0.28	1
0.13	14	0.25	38	**0.25**	**34**	0.3	1
0.2	2	0.28	1	0.28	4	0.31	25
		0.31	10	0.31	33	0.35	21
				0.35	8	**0.4**	**45**
				0.4	13	0.5	1
						0.6	4

DSGE-VAR(5)		DSGE-VAR(6)		DSGE-VAR(7)		DSGE-VAR(8)	
$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency
0.31	4	0.4	11	0.5	2	0.6	19
0.35	4	0.5	9	0.6	32	**0.8**	**72**
0.4	**45**	**0.6**	**41**	0.7	3	1	9
0.5	8	0.8	36	0.8	54		
0.6	28	1	3	1	9		
0.8	11						

APPENDIX A13 MonteCarlo experiment with backward-looking data (Greenspan Sample)

unitary variance shock

DSGE-VAR(1)		DSGE-VAR(2)		DSGE-VAR(3)		DSGE-VAR(4)	
$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency
0.09	**57**	0.17	2	0.24	8	0.3	8
0.1	29	0.2	12	0.25	8	0.31	25
0.13	6	**0.24**	**27**	0.3	29	0.35	17
0.2	2	0.25	25	**0.31**	**37**	0.4	4
0.24	4	0.3	14	0.35	12	**0.5**	**43**
0.25	2	0.31	18	0.4	2	0.6	2
		0.35	2	0.5	4	0.9	1

DSGE-VAR(5)		DSGE-VAR(6)		DSGE-VAR(7)		DSGE-VAR(8)	
$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency	$\widehat{\lambda}$	Frequency
0.35	7	**0.5**	**66**	0.5	21	0.5	1
0.4	3	0.6	5	0.6	9	0.6	5
0.5	**72**	0.7	4	0.7	6	0.7	5
0.6	12	0.9	11	0.9	28	0.9	30
0.9	3	1	14	**1**	**36**	**1**	**59**
1	3						

59