Assessing the Potential of DSGE Model Evaluation in a Bayesian Framework

Alessia Paccagnini[1]

Università Bocconi

March 3, 2009

[1]Ph.D. Student in Economics, E-mail : alessia.paccagnini@unibocconi.it, address: Università Bocconi - Department of Economics - Via Roentgen, 1 - 20136 Milano phone number: +39.02.5836.5443. I am grateful to the invaluable support and guidance from my supervisor Carlo A. Favero. I thank Marco Del Negro for his comments and for having kindly provided me with the Matlab programs used in Del Negro and Schorfheide (2004). I have benefited from helpful discussion of participants of ADDEGeM 2008, Doctoral Meeting of Montpellier (D.M.M.) 1st edition; of participants of First Italian PhD Student Workshop, Collegio Carlo Alberto, Turin and of participants of PhD Student Workshop in Università Bocconi, in particular, Marco Maffezzoli, Bruno Sitzia and Ulf Soderstrom. I would also thank for their valuable comments Elena Besedina, Paolo Bianchi, Agostino Consolo, Claudia Foroni, Riccardo Masolo, Marcella Nicolini, Mario Porqueddu and Giorgio Primiceri.

Abstract

In the recent macroeconometric literature, there has been a growing interest in using Dynamic Stochastic General Equilibrium Models (DSGE) in way of explaining macroeconomic fluctuations and using the models for quantitative policy analysis. Understanding if a certain economic model can explain real data has a prominent place in the research agenda of econometricians and macroeconomists. In two influential papers, Del Negro and Schorfheide (2004) and Del Negro, Schorfheide, Smets and Wouters (2007), an important Bayesian Econometrics procedure to estimate DSGE models by using Vector Autoregressive (VAR) approach is presented. This methodology helps economists to choose the best model to represent real data among a theoretical model, a statistical framework and a combination between the two, a DSGE-VAR representation. The aim of this paper is to study the properties of this famous procedure and to try to highlight some of its aspects carrying out three MonteCarlo experiments which hint the possibility of improvement.

JEL CODES: C11, C15, C32

KEYWORDS: Bayesian Analysis, DSGE Models, Vector Autoregressions

1 Introduction

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

From an econometric point of view, an increasing literature applies Bayesian methods to estimate and evaluate DSGE Models.

Two distinct approaches to macroeconomic analysis have been popular during the early 1980s and continue to be used in the recent research.

First, the standard econometric approach in which an economic model is embedded within a complete probability model and analyzed using classical statistical methods. An example is the use of VAR models, introduced by Sims (1980), 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. Despite of the popularity in their use, VAR models are subject to multicollinearity problems and they can fail to take to account of non-linearities in the economy. Moreover, these statistical models rely so loosely on the economy theory, consequently, 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.

The second approach follows Kydland and Prescott (1982) and attempts to explain the movements and co-movements of many of the same variables using DSGE models. These models draw tight links between the structural parameters describing households' tastes, firms' behaviors, technology and monetary shocks, where these structural parameters should remain invariant to changes in policy regimes. 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).

In order to solve the problems and to combine the advantages of these two approaches, Ireland (2004) proposed the first example of the hybrid model. Following this idea, a prominent new procedure applied to estimate DSGE models by using VAR approach and Bayesian Econometrics is presented by two influential papers, Del Negro and Schorfheide (2004) and Del Negro, Schorfheide, Smets and Wouters (2007a). In this "New Macroeconometrics" literature, the principle underling a Bayesian analysis of DSGE Models[1] is not only to combine prior and likelihood functions to obtain posterior distribution of the interested variables, but also the use of dummy observation priors.

The methodology introduced by Del Negro and Schorfheide (2004) helps economists to choose the best model to represent real data among a theoretical model, a statistical framework and a combination between the two "classical" approaches, a DSGE-VAR representation. However, the idea of using a general equilibrium business cycle model in order to generate a prior for VAR parameters rests upon the Bayesian Econometrics framework is not new in macroeconometrics literature and it has been proposed by Goldberg and Theil (1961) and applied by Doan et al. (1984), Ingram and Whiteman (1994), DeJong et al. (1996 and 2000), Sims (1996). For the first time, in Ingram and Whiteman (1994), it has been shown that shrinking the VAR estimates towards the restrictions implied by a theoretical model (in this case, a neoclassical RBC model, see King et al. (1988)) produced better forecasting performance than one produced using a Minnesota prior (see Doan, Litterman and Sims (1984)).

The advantage of the use of a RBC or a DSGE model prior is the possibility to consider the economic theory which represents the research's beliefs about the mechanisms in the

[1] An important example of evaluation and comparison of DSGE models by using Bayesian Econometrics is presented in Schorfheide (2000).

economy before taking the model to the data. One more relevant point is that the DSGE-VAR model is interpreted as a sort of Structural VAR (SVAR) and under this aspect it is appealing from a policy maker's point of view.

In Del Negro and Schorfheide (2004), there is not only the introduction of this elaborate methodology to create a DSGE-VAR model, but also a procedure to help economists in the choice among the theoretical model [2] , the VAR representation and the hybrid model is presented. The DSGE-VAR model is to be considered as a method for incorporating deviations from the VAR representation of the DSGE model. This method is viewed as a prominent alternative to face the misspecification issue. In the recent literature and in the practice of Central Banks, there exist three different approaches for dealing with misspecification problem. The first approach consists of ignoring the problem and deriving quantitative policy recommendations as if the DSGE model was correctly specified (Laforte (2003) and Levin et al. (2006)). The second approach is to manipulate the shock structure of the DSGE model to optimize the fit of the resulting empirical specification (Smets and Wouters (2003)). The third approach suggests modelling explicitly the deviations from the cross-equation restrictions in the likelihood. For example, Ireland (2004), following Sargent (1989), assumes that the measurement errors improve the empirical fit of a DSGE model, generating serious identification problems and limits in the policy analysis exercises. The DSGE-VAR framework is useful not only to cope with the problem concerning the deviations from the DSGE model but also to assess the robustness of the DSGE model's policy predictions.

On the technical side, DSGE-VAR proposed by Del Negro and Schorfheide (2004) is a mixture model, a combination of an unrestricted VAR for the actual data and a Bayesian VAR implied by the econometric framework of the economic model. The theoretical model

[2] It could be not only an economic model, but any general theoretical model, in this paper the theoretical model considered is a DSGE model.

is treated as a mechanism for generating artificial data (the so-called dummy observation priors, see Sims (2005) for more details), as theoretical second-order moment. These dummy observations represent the restrictions imposed by the econometrician. The degree of the restrictions imposed by the approximation of the DSGE model is governed by a continuous hyperparameter called λ. In Del Negro and Schorfheide (2004) and Del Negro et al. (2007a), this λ represents the weight of the restrictions from the model imposed by the econometrician and it tells how much the economic model (DSGE) is able to explain the real data. When λ is small, the combined model reduces to an unrestricted VAR representation, the real data can be described by using only the statistical framework. When λ approaches ∞, the real data can be explained by using the theoretical model.

The optimal mixture model, DSGE-VAR, is the one associated with the value of λ that maximizes the marginal likelihood for the data, $\widehat{\lambda}$. If $\widehat{\lambda}$ is large, the theoretical model fits the data well, otherwise if $\widehat{\lambda}$ tends to zero, the theoretical model does not describe the data.

Several papers successfully employ this procedure with the aim to compare different DSGE models, moreover for using the DSGE-VAR hybrid model for model validation in what is called the DSGE approach.

For example, in Liu et al. (2008), DSGE-VAR is used to forecast South African Economy, in Adjemian et al. (2008), the hybrid model is used in order to compare different optimal monetary policy and in Adolfson et al. (2008) and Lees et al. (2007), the mixture procedure is used to evaluate open economy models.

Despite of the success of this methodology, some problems and criticisms have been raised in some recent papers[3]. For example, the comments provided by two papers could be considered useful in order to understand better how DSGE-VAR is implemented. The first,

[3] Christiano (2007), Faust (2007), Gallant (2007), Kilian (2007) and Sims (2007).

Christiano (2007) stresses the necessity of furnishing a complete analysis of the marginal likelihood, to assess eventual misspecification in the economic model which is chosen. The second, Kilian (2007) evidences the importance of clarifying the lag length used in VAR representation of the theoretical model. Actually, in the DSGE-VAR combination, the existence and the VAR finite-order representation for the theoretical model (see Ravenna (2007) and Fernandez-Villaverde (2007)) play a key role. Hence, the truncation of the VAR representation has an important impact on the marginal likelihood function. In this procedure, the marginal likelihood and its maximum point depend on the chosen lambda grid. This grid spans from a minimum λ ($\lambda_{\min}$) (necessary in order to get a proper prior) to a maximum λ which represents the maximum weight for the economic model. The minimum λ depends on the degrees of freedom, i.e. the number of observations in the sample size and the number of regressors (endogenous variables multiplied by the lags used in the VAR representation[4]).

Apart of these comments, in the literature there are no papers which propose an empirical analysis of the use of DSGE-VAR approach, discussing what are the main problems and the main advantages. In this lack, there is the main motivation of this paper, which has the purpose to study the properties of this kind of hybrid model, in order to understand its power and its advantages in the recent macroeconometric literature.

The contribution of this paper is twofold. First, this paper discusses and tests the use of DSGE-VAR in the artificial world, trying to point out the main aspects in its use. Second, starting from the results obtained with the artificial data, the paper proposes a robustness analysis of the results provided by Del Negro and Schorfheide (2004).

The method to verify the properties of the hybrid DSGE-VAR in the artificial world is via MonteCarlo experiments. These artificial MonteCarlo experiments are carried out

[4] In this procedure, VAR representation for the actual data has the same number of observations, number of variables and lag length of the Bayesian VAR for the dummy observations.

under the null hypothesis and alternative hypothesis. Under the null hypothesis, the Data Generating Process (DGP) is the model used to generate dummy observation priors in the mixture combination and this model is the candidate one to explain the actual data. Instead, under the alternative hypothesis the DGP is not the model used in the hybrid procedure.

The DSGE-VAR approach is used in order to test these hypotheses. If the optimal lambda is higher than the minimum lambda, with an high percentage, the DSGE model (used to generate dummy observations prior in the procedure) is able to explain the data. Otherwise, if the $\widehat{\lambda}$ is equal to $\lambda_{\min}$, the DSGE model is not able to explain the data, hence, the null hypothesis is rejected.

In the first MonteCarlo experiment, the DGP is the forward-looking model which is the candidate model to explain the data. This model is the same used in the procedure to generate dummy prior observations. The purpose of this experiment is the use of DSGE-VAR approach to verify if the artificial data come really from the model.

In the second MonteCarlo experiment, the artificial data are generated from a backward-looking model (Rudebusch and Svensson (1998), Lindé (2001)) calibrated by using two different parameter sets. Under the null hypothesis, DGP is (as in the previous MonteCarlo) the forward-looking model, but under alternative hypothesis, the DGP is the backward-looking model. Consequently, there is a misspecification in the data generating process.

The purpose of these experiments, by using the backward-looking model, is to understand if the DSGE-VAR has the advantage to recognize if the DGP is the model used to generate dummy observation priors.

As anticipated, the crucial issue in the use of a finite-order VAR representation is the number of lags in the approximation. Following this problem, the hybrid DSGE-VAR is implemented considering a lag-length from 1 to 8. The aim of this implementation is to understand what happens when in the DSGE-VAR the number of lags are misspecified respect to the DGP. As interesting point, it seems that the contribution of the economic model to explain the data increases mispecifying the number of lags in the VAR representation of the theoretical model.

These MonteCarlo experiments are completed by presenting forecasting exercises.

It is obvious that the power of this result depends on the VARMA representation for the economic model, truncated by using a VAR representation (Ravenna, 2007, Fernandez-Villaverde et al., 2007) and this is a compelling aspect in using this methodology.

The use of the artificial world experiment suggests that it is crucial to identify the right number of lags used in VAR representation, since the marginal likelihood depends on the number of degrees of freedom and it seems that the economic model better explains the data in case a VAR representation has too many lags. This problem depends on the fact that the model considers unobservable endogenous variables, such as government and technology processes, as shown in Ravenna (2007). Consequently, the DSGE model has a VARMA representation and it is hard to recognize the finite-order VAR approximation.

Considering this important result, an empirical analysis in the real world is provided, changing the sample size and then the lag length, in order to check what happens to the hybrid DSGE-VAR when more lags are added.

The remainder of the paper is organized as follows. In Section 3.2, 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 3.3, results from MonteCarlo experiments in the artificial world are presented and an empirical analysis in the real world is realized. Concluding remarks are in Section 3.4.

2 The Hybrid Model: a DSGE-VAR Approach

The DSGE-VAR approach, presented in Del Negro and Schorfheide (2004) and discussed in Del Negro, Schorfheide, Smets and Wouters (2007a), uses Bayesian Econometric techniques in order to create a combination between a statistical representation (VAR approach) of the U.S. actual data and a BVAR (based on dummy observation priors) for the artificial data derived by the economic model. The intention of this methodology is to create a hybrid model which combines the characteristics of the data with the characteristics of the candidate economic model to explain data in the estimation, used in this procedure to generate dummy observations (which represent the restrictions given by the model). From a practical point of view, this hybrid model comes from the combination between the likelihood function of the data and the hierarchical prior derived by the parameters in the model. The final result is the posterior that synthesizes all these characteristics.

This procedure is repeated by using different weights (λ) for the economic model.

A numerical optimization procedure is used in order to maximize the marginal likelihood function and this maximum point is the optimal λ, $\widehat{\lambda}$. In the combined model, this optimal value represents how much the economic model explains the data.

First of all, this procedure is presented in a general assessment and hence a simple economic model is used as an example in order to clarify how this method works.

2.1 DSGE-VAR Approach: a general assessment

2.1.1 The Likelihood function

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

$$Y = X\Phi + U \tag{1}$$

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_t.

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

$$L(Y|\Phi, \Sigma_u) \tag{2}$$

2.1.2 Dummy Observation Priors from the Model

First of all, this part regards the theoretical model and the procedure to generate the artificial data. The rational expectations solution of the linearized format of a theoretical model, in this case an economic model, is computed by using the algorithm implemented by Sims (2002).

This solved model can be represented by using the state-space form solution. Adopting the notation in Fernandez-Villaverde et al. (2007):

$$\begin{aligned} x_{t+1} &= A(\theta)x_t + B(\theta)\varepsilon_t \\ y_t &= C(\theta)x_t + D(\theta)\varepsilon_t \end{aligned} \tag{3}$$

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. For simplicity, D is taken as a square and invertible matrix, i.e. the number of shocks is equal to the number of observable variables.

In DSGE-VAR combination, the finite-order VAR truncation to the DSGE model is very important. Fernandez-Villaverde et al. (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:

$$y_t = \sum_{j=1}^{\infty} C \left(A - BD^{-1}C\right)^{j-1} BD^{-1} y_{t-j} + Dw_t \tag{4}$$

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 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 VAR approximation of the economic model is crucial to obtain the prior distribution for the VAR parameters proposed by Del Negro and Schorfheide (2004). 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 :

$$\begin{aligned}
\Phi^*(\theta) &= \Gamma^{*-1}_{xx}(\theta)\,\Gamma^*_{xy}(\theta) \qquad (5)\\
\Sigma^*(\theta) &= \Gamma^*_{yy}(\theta) - \Gamma^*_{yx}(\theta)\,\Gamma^{*-1}_{xx}(\theta)\,\Gamma^*_{xy}(\theta)
\end{aligned}$$

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:

$$Y^* = X^*\Phi(\theta) + U^* \tag{6}$$

where Y^*, X^* and U^* derive from the VAR truncated representation for the theoretical model and coefficients matrix $\Phi(\theta)$ is a function of the parameters used in the model.

In the Del Negro and Schorfheide procedure, these dummy observation priors[5] are assumed to be derived from artificial data based on the simulation of the theoretical model. From the state-space representation of the theoretical model, cross-moments and the likelihood function are computed for the artificial data. This likelihood function results from a flat prior (the Jeffrey prior for the multivariate case) to construct a proper distribution based on the theoretical model. The prior distribution is obtained by fitting the VAR(p) on the data simulated from the structural model, whose length is equal to a fraction, λ, of the length of the actual data.

In this paper, as in Del Negro and Schorfheide (2004), a hierarchical prior is applied. Writing a prior as a hierarchical prior is often a convenient way of expressing prior information.

$$P(\Phi, \Sigma_u, \theta, \lambda) = P\left(\Phi, \Sigma_u | \theta, \lambda\right) P\left(\theta\right) \tag{7}$$

In equation (7), $P\left(\theta\right)$ represents the prior from the model; $P\left(\Phi, \Sigma_u | \theta, \lambda\right)$ is composed by the Bayesian VAR representation for the theoretical model and VAR representation for the real data.

The λ parameter represents the weight of the restrictions imposed on the DSGE model

[5] The Bayesian Econometrics literature suggests the use of dummy observation priors to impose a prior distribution on the set of coefficients. These dummy observation priors have been proposed by Goldberg and Theil (1961) and applied by Ingram and Whiteman (1994) and Sims (1996) (see Sims (2005), for an excellent review). The VAR representation is affected by overfitting problem due to the inclusion of too many lags and too many variables, some of which may be insignificant. The problem of overfitting results in multicollinearity and loss of degree of freedom lead to inefficient estimates.

by the econometrician. When λ approaches ∞, the prior converges to a single spike over $\Phi(\theta)$ and $\Sigma_u(\theta)$ and DSGE model is believed to be true. When λ becomes very small, prior becomes more diffuse and the DSGE model provides very little prior information and if $\lambda = 0$, the prior is useless and the data are represented by only unrestricted vector autoregressive model.

2.1.3 Posterior Distribution and Marginal Likelihood Function

Following Bayes' Rule, the posterior is proportional to the Likelihood times the Prior:

$$P(\Phi, \Sigma_u, \theta | Y, \lambda) \propto L(Y|\Phi, \Sigma_u) P(\Phi, \Sigma_u | \theta, \lambda) P(\theta) \tag{8}$$

In this procedure the prior distribution and the likelihood function are conjugate and the posterior distributions have a typical Normal and Inverted-Wishart format for the coefficient matrix Φ and for the covariance matrix Σ_u. The different combined models are evaluated by different weights for the economic model, λ.

This parameter λ is chosen from a finite grid, $\Lambda = (\lambda_1, \lambda_2, ..., \lambda_\infty)$. The lowest value is 0, and in this case, the best representation for the data is the unrestricted VAR; the second lowest λ is given by the number of parameters, lags, observations in likelihood $\left(\lambda_2 = \frac{n+k}{T}\right)$ in order to get a proper prior density and non-degenerate. The highest λ is ∞, i.e. the data are better fitted by the DSGE model.

In order to evaluate the optimal mixture model, it is useful to clarify the definition of marginal likelihood of the data, conditional on priors and on a specific value of λ:

$$L(Y, \lambda) = \int_{\Phi, \Sigma_u} L(Y|\Phi, \Sigma_u) P(\Phi, \Sigma_u | \theta, \lambda) \, d(\Phi, \Sigma_u) \tag{9}$$

This expression can be evaluated for given values of θ and λ.

The marginal data density is obtained by using an algorithm which allows the econometrician to simulate draws from the distribution $L(Y,\lambda)$ directly without the need to calculate the distribution itself[6].

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

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

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 distance between the data and the model.

2.2 Properties of the Marginal Likelihood Function: an AR(1) example

A simple univariate example (as proposed in Del Negro, Schorfheide, Smets and Wouters (2007a) can be useful to understand better the main properties of the marginal likelihood function $P(Y|\lambda)$. It is possible to consider as a model, the univariate stationary AR(1) process:

$$y_t = \phi y_{t-1} + u_t \qquad u_t \sim N(0,1) \tag{10}$$

The sample autocovariances of order 0 and 1, based on T observations are $\widehat{\gamma_0}$ and $\widehat{\gamma_1}$.

Suppose the theoretical model, for example a DSGE model, represents restrictions on ϕ. Hence, ϕ^* is the corresponding parameter in the model and it represents the restrictions from the theory. The autocovariances of order 0 and 1, implied by the theoretical model, are γ_0 and γ_1.

[6] 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. For more details, see Del Negro and Schorfheide (2004).

By using a DSGE model, the priors are computed, in this simplified case:

$$\phi \sim N\left(\phi^{*}, \frac{1}{\lambda T \gamma_0}\right)$$

where the parameter λ controls the degree of model misspecification with respect, in this case, to the AR(1) process. For small values of λ, the discrepancy between AR and the theoretical model is large; for large values of λ, the discrepancy is small.

In this simple case with a AR(1), the marginal likelihood function of λ takes the following form:

$$\ln p(Y|\lambda, \phi^{*}) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\widetilde{\sigma}^{2}\left(\lambda, \phi^{*}\right) - \frac{1}{2}c\left(\lambda, \phi^{*}\right) \tag{11}$$

The analysis of each component of the marginal likelihood is useful to understand changes in its shape. The first term $\ln(2\pi)$ represents the constant term and it is not related to the economic model.

The second term $\widetilde{\sigma}^{2}\left(\lambda, \phi^{*}\right)$ measures the in-sample one-step-ahead forecast error and it is written as a combination of autocovariances:

$$\widetilde{\sigma}^{2}\left(\lambda, \phi^{*}\right) = \widehat{\gamma_0} + \lambda\gamma_0 - \frac{\left(\widehat{\gamma_1} + \lambda\gamma_1\right)^{2}}{\left(\widehat{\gamma_0} + \lambda\gamma_0\right)} - \lambda\left(\gamma_0 - \frac{\gamma_1^{2}}{\gamma_0}\right)$$

In case $\lambda \to 0$, the $\widetilde{\sigma}^{2}\left(\lambda, \phi^{*}\right)$ converges to the OLS forecast error. Formally speaking,

$$\lim_{\lambda \to 0} \widetilde{\sigma}^{2}\left(\lambda, \phi^{*}\right) = \frac{1}{T}\sum\left(y_t - \widehat{\phi}y_{t-1}\right)^{2}$$

$$\text{where } \widehat{\phi} = \frac{\widehat{\gamma_1}}{\widehat{\gamma_0}}$$

In case $\lambda \to \infty$, the in-sample forecast error is under the restriction implied by the theoretical model. Formally speaking,

$$\lim_{\lambda\to\infty} \widetilde{\sigma}^2(\lambda,\phi^*) = \frac{1}{T}\sum(y_t - \phi^* y_{t-1})^2$$

$$\text{where } \phi^* = \frac{\gamma_1}{\gamma_0}$$

Notice, the in-sample forecast error depends on the length of sample, T.

The general formulation for the forecast error term is monotonically increasing in λ, if λ becomes larger, the forecast error term increases, consequently the marginal likelihood decreases.

The third and last term can be considered as a sort of a "penalty" term for the model complexity.

$$c(\lambda,\phi^*) = \ln\left(1 + \frac{\widehat{\gamma_0}}{\lambda\gamma_0}\right) \tag{12}$$

This penalty is a continuous and decreasing function of the hyperparameter λ. If λ approaches ∞, there is no parameter to be estimated in AR process and the model has importance. If λ goes to 0, the autoregressive parameter ϕ is completely unrestricted. Usually, in the "classical" selection problem, the penalty term is an increasing function of the number of included regressors. It is useful to remember that λ is chosen from a lambda grid which spans from a minimum value of λ which depends on the number of observations in the sample size, the number of the endogenous variables and the number of lags. Consequently, the bigger $\lambda_{\min}$ depends on the more complexity of the model.

However, this simple example is very useful to stress out three main properties of the marginal likelihood function.

First, if the sample autocovariances are very different from the autocovariances derived under the restriction $\phi = \phi^*$, the marginal likelihood peaks at a small value of λ. When the discrepancy between the sample and the DSGE model autocovariances decreases, the optimal λ, increases and the marginal likelihood will attain its optimal $\widehat{\lambda} = \infty$, i.e. the theoretical model represents perfectly the real data.

Second, if the parameter $\lambda \to 0$, the marginal likelihood function tends to minus infinity. It is obvious that in case of a multivariate statistical framework VAR, this feature of the marginal likelihood function enforces parsimony and prevents the use of over-parameterized specifications that cannot be precisely estimated based on the fairly short samples.

Lastly, it is possible to compare two different models. Suppose two models M_1 and M_2 with two different priors ϕ_1^* and ϕ_2^*. For small values of λ the goodness-of-fit term $\widetilde{\sigma}^2 (\lambda, \phi_1^*)$ and $\widetilde{\sigma}^2 (\lambda, \phi_2^*)$ are essentially identical and the only difference in the log likelihood are given by the differences in the penalty terms. For large values of λ, penalty differentials become less important and the marginal likelihood comparison is driven by the relative in-sample fit of the two restricted specifications. If autocovariances associated with M_1 are closer to the sample autocovariances than autocovariances associated with M_2, then $\widehat{\lambda_1} > \widehat{\lambda_2}$. However, it is important to evidence that the penalty term (12) does not depend directly on the variables of the AR representation and on the priors of the model, but only on the hyperparameter λ and on the ratio between sample autocovariance and theoretical autocovariance.

Concerning the general properties of the marginal likelihood in this DSGE-VAR application, there are several useful comments in the recent literature (Christiano (2007), Faust (2007), Gallant (2007), Kilian (2007) and Sims (2007)).

First of all, Christiano (2007) suggests the necessity of a further analysis of the shape of the marginal likelihood in order to apprehend if a DSGE model, in this case, could be considered a good model. It is obvious that the marginal likelihood is influenced by priors over the model parameters[7].

Christiano proposes two MonteCarlo experiments in which artificial data are generated by a DSGE model and the econometrician correctly specifies the model. If $\widehat{\lambda}$ is small for the econometrician, there is something wrong with the DSGE model. In another MonteCarlo experiment, Christiano employs the econometrician's DSGE model misspecified. The determining result found by these experiments is that DSGE-VAR approach is useful to identify weakness in model fit; moreover, an analysis about the rate at which the marginal likelihood declines for $\lambda > \widehat{\lambda}$ could be useful. Christiano's MonteCarlo experiments evidence that a steep rate of decline is a signal that the econometrician's DSGE model fits poorly and Christiano suggests to provide a measure for this rate; for example, reporting Bayesian probability intervals for the hyperparameter λ.

Alternatively, a forecasting evaluation could be taken into consideration to compare DSGE-VAR with Bayesian VAR and unrestricted VAR.

Del Negro et al. (2007b), reply to Christiano's comments, explaining that λ could be not considered as a classical test of the hypothesis in which DSGE model restrictions are satisfied and they stress the only Bayesian interpretation for the marginal likelihood function $L(Y, \lambda)$; consequently, it is not necessary any cutoff or critical values but it could be important to study the entire shape of the marginal likelihood function.

Second, in Kilian (2007), the main point of the discussion is the lag length used in VAR estimation for the DSGE-VAR and moreover the lack of reasons provided by Del Negro et al. (2007a) concerning the choice to use a certain lag length in VAR representation.

Del Negro et al. (2007b) explain that there are essentially two dimensions to the choice of the lag order in a DSGE-VAR. The first dimension is the empirical fit of the DSGE-VAR with the optimal value of λ , that is the DSGE-VAR($\widehat{\lambda}$). This suggests that we choose the

[7] As Chris Sims explains in an interview realized by Hansen (2004) the theoretical model is important in order to generate priors, relevant for forecasting purposes.

lag-length to maximize the marginal data density associated with the DSGE-VAR($\widehat{\lambda}$).

The second dimension of the lag-length choice is related to the accuracy of the VAR approximation to the DSGE model. The lag-length has been chosen to minimize the approximation error that is to minimize the discrepancy between the dynamics of the model DSGE-VAR(∞) and the dynamics of the DSGE model. However, the accuracy of the approximation increases with lag length, this criterion leads to take a large number of lags in the approximation.

In any case, they do not provide a theoretical proof in order to explain this aspect and they use four lags in the VAR representation. However, according to An and Schorfheide (2007), when quarterly real time series are used in the empirical analysis, a VAR with four lags can be considered as a "good" truncation of the VARMA representation.

Considering all these aspects, it is possible to point out that the problem of the misspecification of the lags in the VAR representation has an important role in the composition of DSGE-VAR. The misspecification can lead an increase in the marginal likelihood and in the marginal data density associated to the optimal lambda, $\widehat{\lambda}$. In this sense, it is possible to focus on the penalty term discussed above (12) compared to the penalty term in the usual Information Criteria.

2.3 Properties of the Marginal Likelihood Function: the Penalty term

In order to comment the penalty term, it could be useful to analyze a more detailed generic representation of the marginal likelihood. Taking the general Laplace expansion for a generic Marginal Likelihood function (Robert, 2007):

$$\int_{\Theta} \exp\left\{nh\left(\theta\right)\right\} d\theta = \exp\left\{nh\left(\theta\right)\right\} \left(2\pi\right)^{p/2} n^{-p/2} \left|H^{-1}\left(\widehat{\theta}\right)\right| + O(n^{-1}) \tag{13}$$

where Θ is the parameter space associated with the set of the models, p is the dimension of Θ, $\widehat{\theta}$ is the maximum of h and H is the Hessian of h. In model selection theory, the Posterior Odds is decomposed into Priors Odds and Bayes Factor. Suppose there are two models, M_1 and M_2 to be compared:

$$B_{12}^{\pi} \simeq \frac{L_{1,n}\left(\widehat{\theta}_{1,n}\right)}{L_{2,n}\left(\widehat{\theta}_{2,n}\right)} \left|\frac{H^{-1}\left(\widehat{\theta}_{1,n}\right)}{H^{-1}\left(\widehat{\theta}_{2,n}\right)}\right|^{1/2} \left(\frac{n}{2\pi}\right)^{(p_2-p_1)/2} \tag{14}$$

where p_1 and p_2 are the dimensions of Θ_1 and Θ_2, $L_{1,n}$ and $L_{2,n}$ are the likelihood functions based on n observations, and $\widehat{\theta}_{1,n}$ and $\widehat{\theta}_{2,n}$ are the maxima of L_1 and L_2, respectively.

This Bayes Factor is decomposed into two parts, the Likelihood ratio and the Occam Factor.

The Likelihood ratio is :

$$\frac{L_{1,n}\left(\widehat{\theta}_{1,n}\right)}{L_{2,n}\left(\widehat{\theta}_{2,n}\right)} \tag{15}$$

The other terms are what is called in the theory, the Occam Factor (MacKay, 2003). The main purpose of the Occam Factor is in the comparison models. Models with more parameters are usually able to provide the best fit for the data. Hence, the Occam Factor penalizes models for "wasted" volume of parameter space.

However, the Occam Factor has several problems. First of all, this factor depends on the prior and the prior should be proper. If the econometrician considers two identical models with different priors, the factor suggests that the model with the best fitting prior has a big evidence. Hence, the choice of prior range is very important and critical.

Following the equation (14), the Bayes Factor could be transformed as:

$$\log(B_{12}^{\pi}) \simeq \log \omega_n + \frac{p_2 - p_1}{2} \log(n) + K(\widehat{\theta}_{1,n}, \widehat{\theta}_{2,n}) \tag{16}$$

where ω_n is the standard likelihood ratio for the comparison of M_1 with M_2

$$\omega_n = \frac{L_{1,n}\left(\widehat{\theta}_{1,n}\right)}{L_{2,n}\left(\widehat{\theta}_{2,n}\right)} \tag{17}$$

From this approximation of the Bayes Factor, the Schwarz's criterion (1978) is derived as:

$$S = -\log \omega_n - \frac{p_2 - p_1}{2} \log(n) \tag{18}$$

when $M_1 \subset M_2$, if the reminder term $K(\widehat{\theta}_{1,n}, \widehat{\theta}_{2,n})$ is negligible compared with both other terms, that is, is a $O(1)$.

For regular models, when $M_1 \subset M_2$, the likelihood ratio is approximately distributed as a $\chi^2_{p_2-p_1}$ distribution,

$$-2 \log \omega_n \approx \chi^2_{p_2-p_1}$$

if M_1 is the true model.

The Schwarz's criterion, also called BIC (Bayes Information Criterion) provides a first-order approximation to the Bayes factor (Kass and Raftery, 1995). This Information Criterion has several problems in a Bayesian setting. First of all, the dependence on the prior assumption disappears and it is obvious that the approximation only works for regular models.

Spiegelhalter, Best and Carlin (1998) and Spiegelhalter, Best, Carlin and Van Der Linde (2002) have developed a hierarchical modeling generalization of AIC (Akaike's Information criterion) and BIC, based on the deviance and called DIC (Deviance Information criterion). This criterion is more satisfactory than two former alternatives because it takes into account the prior information and gives a natural penalization factor to the log-likelihood. Moreover, this Bayesian criterion is particularly useful in model selection problems where the posterior distributions of the models have been obtained by MonteCarlo Markov Chain simulation. This criterion is an asymptotic approximation (like AIC and SIC) as the sample size becomes large. Consequently, it is only valid when the posterior distribution is approximately a multivariate normal. Besides, it also allows for improper priors, since each model is considered separately.

It is possible to see in some details, how this Information Criterion is derived. The starting point is a model $f(x|\vartheta)$ which is associated with a prior distribution $\pi(\vartheta)$, the deviance is defined as follows:

$$D(\theta) = -2\log(f(x|\theta)) + C$$

where x are the data, θ are the unknown parameters of the model and $(f(x|\theta))$ is the likelihood function. C is a constant that cancels out in all calculations that compare different models, and which therefore does not need to be known. The deviance as defined as earlier is not a good discriminating measure, given its bias toward higher dimensional models.

It is possible to define the measure of how the model fits the data as:

$$\overline{D} = E[D(\theta)|x]$$

The larger this measure is, the worse the fit.

The effective number of parameters of the model is computed as:

$$p_D = \overline{D} - D(\overline{\theta})$$

where $\overline{\theta}$ is the expectation of θ. The larger this is, the easier it is for the model to fit the data.

The Deviance Information Criterion could be expressed as follow:

$$\begin{aligned} DIC &= E[D(\theta)|x] + p_D \\ &= E[D(\theta)|x] + \{E[D(\theta)|x] - D\left(E[\theta|x]\right)\} \end{aligned}$$

The factor $E[D(\theta)|x]$ can be interpreted as a measure of fit while p_D is a measure of complexity, also called the effective number of parameters[8].

The idea is that models with smaller DIC should be preferred to models with larger DIC. Models are penalized both by the value of $\overline{D}$, which favors a good fit, but also by the effective number of parameters. $\overline{D}$ decreases as the number of parameters in a model increases, the p_D term compensates for this effect by favoring models with a smaller number of parameters. The main advantage of DIC over other criteria, for Bayesian model selection, is that it is easily calculated from the samples generated by a Markov Chain MonteCarlo simulation. AIC and BIC require calculating the likelihood at its maximum over θ, which is not readily available from MCMC simulation. This information criterion could be useful to take into consideration problems concerning the priors.

This discussion on the classical Information Criteria is useful to stress that the penalty term used in the marginal likelihood function of the hybrid model, DSGE-VAR, is not able to recognize the problem of the misspecification of the number of the parameters estimated (in this case, the number of lags). Comparing the penalty term of the marginal likelihood function to the penalty term of Deviance Information Criterion, it is also possible

[8] Since DIC=$D[E(\theta)|x]+2p_D$, the analogy with AIC is clear. As shown in Spiegelhalter, Best and Carlin (1998), in a non-hierarchical framework where the posterior distribution of θ is approximately normal, DIC and AIC are equivalent.

to notice that the priors should be important. According to DIC, the priors are relevant in the comparison of mixture models and the penalty term (12) of the marginal likelihood function of DSGE-VAR should be formulated in a way more precise.

In this paper, the classical Information Criteria (AIC, SIC and HQ (Hannah-Quinn)) are used in the exercises of model selection in order to recognize the best statistical representation. In these applications, the priors influence is ignored. However, following Del Negro et al. (2007b), another criterion is implemented. The overall DSGE-VAR combination is considered in a maximization of the marginal data density exercise. In this way, it is possible to check what is the lag allowing the econometrician the maximum value of the marginal data density, combining both the actual data and the artificial data.

2.4 DSGE-VAR Approach: an example

A simple economic model can be used as an example of theoretical model. The possible candidate model is a forward-looking model, a simple small-scale New-Keynesian model.

This model is the candidate to explain the actual U.S. time series for real GDP, CPI and Federal Funds Rate over the period 1981:1-2001:4[9].

From the trivariate VAR representation of the data, the likelihood function is the following (as presented in the previous section):

$$L(Y|\Phi, \Sigma_u) \tag{19}$$

[9]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.

The second step is to consider the dummy observation priors which come from the theoretical model.

The economy described in the theoretical model 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, there is a perfectly competitive, representative final goods producer which uses 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. The nominal rigidities are introduced in terms of price adjustment costs for the monopolistic firms. It is obvious that each firm maximizes the profits over infinite lifetime by choosing labour input and its price.

The third component in this economy is the government. This authority 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:

$$\widetilde{x}_t = E_t[\widetilde{x_{t+1}}] - \frac{1}{\tau}(\widetilde{R}_t - E_t[\widetilde{\pi_{t+1}}]) + (1-\rho_g)\widetilde{g}_t + \rho_Z \frac{1}{\tau}\widetilde{z}_t \tag{20}$$

$$\widetilde{\pi}_t = \beta E_t[\widetilde{\pi_{t+1}}] + \kappa[\widetilde{x}_t - \widetilde{g}_t] \tag{21}$$

$$\widetilde{R}_t = \rho_R \widetilde{R_{t-1}} + (1-\rho_R)(\psi_1 \widetilde{\pi}_t + \psi_2 \widetilde{x}_t) + \epsilon_{R,t} \tag{22}$$

$$\widetilde{g}_t = \rho_g \widetilde{g_{t-1}} + \epsilon_{g,t} \tag{23}$$

$$\widetilde{z}_t = \rho_z \widetilde{z_{t-1}} + \epsilon_{z,t} \tag{24}$$

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 first equation 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 second equation 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 third equation 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 an exact finite order representation. A VARMA representation is needed in the presence of these unobservable variables.

The rational expectations solution of the linearized model is then computed using the algorithm implemented by Sims (2002). The first step towards solution is to cast the model in the following form :

$$\Gamma_0 \widetilde{\mathbf{Z}}_t = \Gamma_1 \widetilde{\mathbf{Z}}_{t-1} + C + \Psi \epsilon_t + \Pi \eta_t \tag{25}$$

$t = 1, ..., T$ 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 $E_t\left(\eta_{t+1}\right) = 0$, all t. The results are as follows:

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

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

$$\Gamma_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \rho_R & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho_g & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \rho_z & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \Psi = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \Pi = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

As a solution the following policy function that represents the transition equation is obtained:

$$\widetilde{\mathbf{Z}}_t = \mathbf{T}(\theta)\,\widetilde{\mathbf{Z}}_{t-1} + \mathbf{R}(\theta)\,\varepsilon_t \tag{26}$$

$$\theta = [\kappa, \tau, \psi_1, \psi_2, \rho_R, \rho_g, \rho_z, \sigma_R, \sigma_g, \sigma_Z]' \tag{27}$$

This transition equation 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:

$$\mathbf{Z}_t = W(\theta)\widetilde{\mathbf{Z}}_t + D(\theta) + \upsilon_t \tag{28}$$

where

$$\mathbf{Z}_t = \begin{bmatrix} \Delta \ln x_t \\ \Delta \ln P_t \\ \ln R_t^a \end{bmatrix} \tag{29}$$

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

$$\begin{aligned} \Delta \ln x_t &= \ln \gamma + \Delta \widetilde{x}_t + \widetilde{z}_t \qquad (30) \\ \Delta \ln P_t &= \ln \pi^* + \widetilde{\pi}_t \\ \ln R_t^a &= 4[(\ln r^* + \ln \pi^*) + \widetilde{R_t}] \end{aligned}$$

The economic model is characterized by the following set of parameters:

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

As introduced above, the DSGE model imposes tight restrictions across the parameters of the moving average (MA) representation for output growth, inflation and interest rates. The economic model is written using a VARMA representation that can be very closely approximated by a finite order VAR representation (Ravenna, 2007, Fernandez-Villaverde et al., 2007), Del Negro and Schorfheide (2004) propose to evaluate the DSGE model by assessing the validity of the restrictions imposed by such model on an unrestricted VAR representation for the vector of the three variables. They use a truncation with four lags length.

Consequently, as specified in the previous section, the theoretical model is represented in the following VAR format:

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

where Y^*, X^* and U^* are the statistical representation for the economic model and $\Phi(\theta)$ is in function of the vector of parameters for the candidate model.

As aforementioned, the hierarchical prior is:

$$P(\Phi, \Sigma_u, \theta, \lambda) = P\left(\Phi, \Sigma_u | \theta, \lambda\right) P\left(\theta\right)$$

where in this particular case, the prior $P(\theta)$ is specified as follows:

TABLE 1. 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 marginal likelihood function of the parameter vector, θ is given by:

$$L(Y,\lambda) = \int_{\Phi,\Sigma_u} L(Y|\Phi,\Sigma_u) P\left(\Phi,\Sigma_u|\theta,\lambda\right) d\left(\Phi,\Sigma_u\right)$$

The shape of the marginal data density depends on several factors: the priors on the parameters of the chosen theoretical model, the length of the sample of the data and the number of regressors in VAR representation, especially the lag-order of the VAR approximation considered.

By taking into account the sample size and the number of the regressors, the minimum value for λ in the lambda grid is computed in the following way:

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

Remember that $\widehat{\lambda} \geq \lambda_{MIN}$ in order to get a prior density and non-degenerate, which is a necessary condition for computing meaningful marginal likelihoods.

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

In this paper, the ratio $\frac{\widehat{\lambda}-\lambda_{MIN}}{\lambda_{MIN}}$ is considered as the measure to understand if the DSGE model can explain the real data. This ratio evidences how much in the DSGE-VAR, the economic model explains the actual data over the statistical framework.

3 The empirical analysis

The main motivation of this paper is to provide a study of the properties of the hybrid model in order to understand the power of the use of this new econometric procedure applied to DSGE models.

The paper presents two MonteCarlo experiments in order to assess DSGE-VAR procedure. These artificial MonteCarlo experiments are carried out under the null hypothesis and alternative hypothesis. Under the null hypothesis, the Data Generating Process (DGP) is the model used to generate dummy observation priors in the mixture combination and this model is the candidate one to explain the actual data. Instead, under the alternative hypothesis the DGP is not the model used in the hybrid procedure.

The DSGE-VAR approach is used in order to test these hypotheses. If the optimal lambda is higher than the minimum lambda, with an high percentage, the DSGE model (used to generate dummy observations prior in the procedure) is able to explain the data. Otherwise, if the $\widehat{\lambda}$ is equal to $\lambda_{\min}$, the DSGE model is not able to explain the data, hence, the null hypothesis is rejected.

The first MonteCarlo experiment consists of generating data from the forward-looking model, which is the candidate model used to explain the data and it is the model used to generate dummy observation priors in the DSGE-VAR approach. The aim of this experiment is to evaluate if these artificial data are really described by the candidate economic model.

In the second MonteCarlo experiment, under the alternative hypothesis, the DGP is a backward-looking model (Rudebusch and Svensson (1998), Lindé (2001)), calibrated by using two different parameter sets.

The purpose of this kind of the experiments is to use these artificial data coming from a model which is not the estimated model for the data.

The intention of this exercise in the artificial world is not only to understand what happens in case of misspecification of the DGP, but also when there is a misspecification of the number of lags in DSGE-VAR. After these experiments, a further analysis is realized on real data to understand the problems occur mispecifying the number of lags in VAR representation of the hybrid model.

3.1 MonteCarlo design: Generating data from a forward-looking model

The first MonteCarlo experiment is carried out generating artificial data which comes from the forward-looking model used to generate dummy observation priors and it is the candidate to explain the actual data.

In order to generate artificial data[10], the VARMA representation of the economic model is taken into consideration. This representation of the model is given by the Sims' algorithm.

The number of the generated artificial series is the same as the number of the states of state-space representation of the economic model. The first three states refer to Real GDP, CPI and Federal Funds Rate, hence the three first artificial time series related to these three states are chosen to be the artificial data to be estimated. These series are generated for 80 observations which represent the small sample size from 1981 to 2001 (the same sample used in Del Negro and Schorfheide (2004)).

The MonteCarlo experiment is carried out and DSGE-VAR approach to compute the optimal λ, is used. This procedure is replicated taking into consideration a grid of possible lags, from one to eight, in VAR representation. It is important to stress out that the artificial data coming from the economic model can be represented in a VAR truncated at the first-lag. The purpose of this experiment is to understand the behavior of DSGE-VAR in case of misspecification of the number of lags used in the VAR representation. The minimum λ [11] for each case of VAR representation, considering the different lags, are included in the lambda grid[12]. Artificial data are generated by a MonteCarlo experiment with 100 replications[13].

[10] 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 1.

[11] 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).

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

The bigger lambda considered in this set is 1, in this case the exercise does not aim to draw the shape of the marginal likelihood and it is preferable to take into consideration a more compact lambda around the possible optimal lambda.

[13] In this case, the number of replications in the algorithm to compute the posterior density are 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.

The following table summarizes the frequency of the optimal λ for each different VAR representation (between the brackets (DSGE-VAR()) the number of lags are indicated):

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 2. 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				

The data generating process is a VAR(1) and in case of only one lag in DSGE-VAR the $\widehat{\lambda}$ is 0.09 with a percentage of 27% and it is equal to 0.10 with a percentage of 43%. In this exercise, in 73% cases, the optimal lambda, $\widehat{\lambda}$, is greater than the minimum lambda. This result suggests that there is a contribution of the economic model to explain the actual data[14].

Adding lags to DSGE-VAR increases the optimal lambda, $\widehat{\lambda}$, which becomes greater than the minimum lambda. For example, in DSGE-VAR with 3 lags the minimum lambda is 0.17, the optimal lambda grid does not contain the minimum lambda and in 67% cases the $\widehat{\lambda}$ is equal or greater than 0.28. The same pattern emerges when adding more lags. This result is relevant since it tells the econometrician that by adding more lags, and thus increasing the $\widehat{\lambda}$, the economic model gets more weight than the statistical representation; even if there is a misspecification in the number of lags.

Consequently, the null hypothesis is not rejected, the Data Generating Process come from the forward-looking model and this result becomes relevant mispecifying the lag length.

In these MonteCarlo experiments, usual Information Criteria are used to assess the lag-length. In this sense, the maximization of the marginal likelihood for lag-length purposes is not used as suggested by Del Negro at al. (2007b). In the Appendix, a table shows how in one replication of the MonteCarlo experiment, with VAR(1) as DGP, the maximization of the marginal likelihood leads to choose one lag, as it could be suggested by the usual Information Criteria. Hence, it is possible to use the Classical Information Criteria without problems.

[14] The fact that in 70%, $\widehat{\lambda}$ is equal to 0.09 or 0.10 could be misleading and it seems that there is not a so big contribution of the model to explain the data. However, this point could depend on the small sample of 80 observations of the DGP.

In the Appendix, there is a table in which 1000 observations sample are considered in one MonteCarlo replication.

In the next table, the usual Information Criteria [15] are compared to the maximum among the optimal lambda, $\widehat{\lambda}$, found and to the shape of the ratio. As explained before, the ratio is given by $\frac{\widehat{\lambda}-\lambda_{MIN}}{\lambda_{MIN}}$ and it represents how much the economic model is able to to explain the data over the pure statistical representation.

TABLE 3. Summary Table

AIC		SIC		HQ		max 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**		

First of all, the usual Information Criteria suggest the data are represented by a VAR with only one lag, as the DGP. The maximum lambda is the maximum value assumed by lambda among the $\widehat{\lambda}$ across the different lags for the same artificial data, in each MonteCarlo experiment. With a percentage of 57% this maximum lambda is at lag 8, it means the $\widehat{\lambda}$ is increasing across the lags for the same artificial data.

The shape of the ratio suggests that in only 24% of all cases the ratio decreases after the first lag. Hence, this ratio is increasing after the first lag with a percentage of 76%.

This result suggests an increasing ratio, hence the economic model contribution is increasing in number of lags. In this sense, it is possible to find a link between the shape of this ratio and the Information Criteria, an opposite relationship.

[15] Akaike, AIC, Schwarz, SIC and Hannan-Quinn, HQ.

However, this result is very impressive since artificial data are generated as a VAR with only one lag and DSGE-VAR procedure shows how adding more lags in the statistical representation, the optimal λ is bigger, but the main reason is that the lambda grid obviously changes, consequently the ratio that summarizes the explanation of the theoretical model could increase. This aspect suggests that adding more lags enables the economic model to explain better the data and the procedure could have misspecification problems.

3.1.1 Forecasting

The MonteCarlo analysis is completed by considering the out-of-sample forecasting performance of VAR, DSGE and DSGE-VAR models. All 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 the table, 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.

TABLE 4. Forecasting

	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.

3.2 MonteCarlo design: Generating Data from a Backward-Looking Model

In the second MonteCarlo experiment, under the alternative hypothesis, the DGP come from a different model from the candidate model used in the DSGE-VAR combination, a backward-looking model (Rudebusch and Svensson (1998), Lindé (2001)), calibrated by using two different parameter sets.

The purpose of this kind of the experiments is to use these artificial data coming from a model which is not the estimated model for the data and check if the DSGE-VAR approach is able to recognize this misspecification.

The Rudebusch and Svensson model, which is drawn on the theoretical model presented by Svensson (1997) could be considered a good approximation of real data. It 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.

This model consists of AS and AD equations relating to the output gap (y) (the percentage deviation of output from its steady state level), to the inflation rate (π) and to the monetary policy instrument, the short-term interest rate (i).

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

$$\pi_t = \alpha_{\pi_1}\pi_{t-1} + \alpha_{\pi_2}\pi_{t-2} + \alpha_{\pi_3}\pi_{t-3} + \alpha_{\pi_4}\pi_{t-4} + \alpha_y y_{t-1} + \varepsilon_t^{\pi} \tag{31}$$

$$y_t = \beta_{y_1} y_{t-1} + \beta_{y_2} y_{t-2} + \beta_r \sum_{j=1}^{4} \frac{1}{4} (i-\pi)_{t-j} + \varepsilon_t^{y} \tag{32}$$

$$i_t = \gamma i_{t-1} + \varepsilon_t^{i} \tag{33}$$

In the AS equation, (31), the annualized inflation rate π depends: on past inflation rates, the output gap in the previous period and an exogenous supply shock ε_t^{π} (i.i.d. with zero mean and constant variance σ_{π}^2).

In the AD equation, (32), output gap y_t is related to past output gaps y_{t-1} and y_{t-2}, 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, (33), 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 changes (Burns, Volcker, Greenspan).

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

The following table presents the estimated coefficients proposed by Rudebusch and Svensson and Lindé.

TABLE 5. 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

As aforementioned, the estimation of the parameters used in generation of artificial data comes from Lindé experiments. In this paper, by using the backward-looking model as DGP, there two MonteCarlo exercises. 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. The period in which Greenspan led FED is around the same of the 80 quarters considered in VAR estimation in Del Negro and Schorfheide (2004). As regards the monetary policy process, there is no any indication about the calibration of the autoregressive coefficient. This coefficient has been estimated on the Federal Funds Rate time series for both the whole sample and Greenspan sample and it is around 0.9. Instead, the standard error of the monetary policy shock has been estimated 1.34 for the whole sample and 0.69 for Greenspan sample.

In these MonteCarlo experiments, the econometrician expects to reject the null hypothesis by using the DSGE-VAR approach. The aim of these exercises is to understand if the hybrid model approach is useful to recognize that the DGP is not the model used to generate dummy observation priors. Another crucial point is to understand, by using the DSGE-VAR procedure, what happens misspecifying the number of lags in the VAR representation.

This backward-looking model has a convenient state-space representation:

$$X_t = AX_{t-1} + \upsilon_t \tag{34}$$

where

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

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

The three shocks are distributed as a standardize normal (with zero as mean and one as variance). The artificial data are generated by using a VAR representation with three lags.

In the first MonteCarlo experiment, the 80 observations artificial data are generated by using the parameters and variance estimated values for the whole sample from 1970Q1 to 1997Q4. The null hypothesis is the data come from the forward-looking model, the econometrician expects to reject the null hypothesis. In this experiment, the lambda grid chosen is very similar to the set used in the last section for the same motivations explained in the previous MonteCarlo experiment[16]. The artificial data are generated by a MonteCarlo experiment with 100 replications[17].

The following table has the same structure of the table presented in the previous experiment.

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

[17] In this case, the number of replications in the Metropolis-Hastings algorithm to compute the posterior density are 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 6. 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 exercise, the Data Generating Process is a VAR with 3 lags. The artificial data come from a different model than the model used in the DSGE-VAR composition. The expected result is to get the optimal lambda equal to the minimum lambda. In every DSGE-VAR exercise, the minimum lambda is also among the possible $\widehat{\lambda}$ and it seems 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%.

Hence, the sample autocovariances are very different from the autocovariances derived under the restriction on the cross-moments, as a result is the marginal likelihood peaks at a small value of λ.

The DSGE-VAR recognizes that the DGP is not the model used to generate dummy observation priors.

TABLE 7. Summary Table

AIC		SIC		HQ		max 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								

The statistical representation of this model is a VAR with 3 lags. The maximum value for lambda is reached at lag 8 in 94% of cases, it means that $\widehat{\lambda}$ is increasing across the lags for the same artificial data and the ratio $\frac{\widehat{\lambda}-\lambda_{MIN}}{\lambda_{MIN}}$ falls after the first lag with a percentage of 31%. In this case the ratio is not helpful to stress the number of lags, it tells the econometrician the ratio falls the first time at lag 5 with a percentage of 40%, but in the majority of the cases this ratio is equal zero until the lag 5, since $\lambda_{\min} = \widehat{\lambda}$. This exercise helps to understand that if data come from a backward-looking model, DSGE-VAR procedure is able to evidence this problem and consequently it is possible to reject the null hypothesis that the data come from a forward-looking model.

However, the artificial data have been generated by considering an estimation for the parameters and for the variance of the shocks which has been realized on a sample from

1960 to 1997. In this long period, U.S. economy has faced several monetary policy regimes and crises. The estimation could be affected by this problem, hence it is possible that an analysis based on a small sample could be better. Actually, it is possible to implement the same procedure by considering the estimation of coefficients for the period where Greenspan has been the Chairman of FED (1987Q3 to 1997Q2). The point is that the priors used to generate the artificial data in the DSGE-VAR combination refer to the period 1981-2001 (see, Del Negro and Schorfheide (2004)). Hence, it could be better use the calibration of the model parameters for the Greenspan sample.

As before, by using MonteCarlo experiment, the artificial data with 80 observations are generated 100 times from the backward-looking model[18]. The Data Generating Process is a VAR representation with three lags.

In the following table, which has the same structure of the table presented in the previous experiments, it is shown that adding lags to VAR increases the optimal λ.

[18] In this case, the number of replications in the algorithm to compute the posterior density are 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 8. 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						

In this exercise, the artificial data come from a different model than the model used in DSGE-VAR composition. Consequently, the null hypothesis that Data Generating Process comes from the forward-looking model is expected to be rejected. The expected result is to get the $\widehat{\lambda}$ equal to the minimum lambda. 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. In case of DSGE-VAR with 3 lags, only in 33% cases the minimum lambda is equal to the maximum lambda. Adding lags makes the $\widehat{\lambda}$ bigger than the minimum lambda. Consequently, the misspecification of the number of the lags gives more weight to the economic model.

In the next table, the usual Information Criteria [19] are compared to the maximum among the optimal lambda found and to the shape of the ratio.

TABLE 9. Summary Table

AIC		SIC		HQ		max 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								

The statistical representation of this model is a VAR with 3 lags. The maximum value for lambda is reached at lag 8, it means that $\widehat{\lambda}$ is increasing across the lags for the same

[19] Akaike, AIC, Schwarz, SIC and Hannan-Quinn, HQ.

artificial data and the ratio increases after the first lag in 69% of cases. In this exercise there are two kinds of misspecification, first, the DGP is not the forward-looking model and second, the number of lags. Adding lags leads to not reject the null hypothesis, hence the DSGE-VAR is not able to recognize the misspecification.

In this sense, the use of classical Information Criteria could help the economist to avoid these problems. Since it seems the priors of the parameters of the model could be important in this discussion, the use of DIC can be useful to overcome this point.

3.2.1 Forecasting

The forecasting performance concludes these MonteCarlo experiments. In this first table, the forecasting evaluation in MonteCarlo experiment is presented when parameters of the backward-looking model are calibrated by using the whole sample calibration.

TABLE 10. Forecasting (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 11. Forecasting (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.

3.3 Comments on Results

The two different MonteCarlo Experiments are useful to stress out some important points. First of all, when data come from the forward-looking model, the DSGE-VAR recognizes that DGP is the same model used to generate dummy observation priors

Adding lags seems to make the economic model more important in explaining the data. But the data come from this model and this result might be not so surprising. The misspecification gives more weight to the economic model over the statistical framework.

When generating data from a different model that one used to generate dummy observations in the DSGE-VAR approach, the $\widehat{\lambda}$ is very near to the minimum lambda and this result is expected since the forward-looking model should not matter when these data are used. Using Greenspan sample calibration, this result is not always true, the optimal lambda is greater than the minimum lambda, mispecifying the number of lags. In this case, the misspecification leads to a more contribution of the economic model.

The crucial point as evidenced by Ravenna (2007) is the truncation of the VARMA representation of the theoretical model. Additional lags are not penalized in the marginal likelihood which increases with the misspecification of the number of lags. However, this problem is relevant and important when the economist works in the real world and she does not know the DGP.

It is obvious that these results depend on the number of replications used in the Metropolis-Hastings and it is possible to get an accurate result, increasing the number of the replications, but this exercise is very time-consuming. Moreover, it is possible to change the variance shocks in the backward-looking model. Actually, when the econometrician takes into consideration a unit variance for shock, the result changes drastically. The economic model becomes more important in both sample 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 in Appendix.

3.4 Empirical Results in the Real World

The MonteCarlo experiments show that the DSGE-VAR can recognize the Data Generating Process, when there is a misspecification of lags, there is more contribution of the economic model. In this section, an empirical analysis in the real world is carried out. The aim is to apply the DSGE-VAR approach in the real world, focusing on the sample size and on the number of lags.

In the previous sections, it has been discussed that the lambda set depends on the number of endogenous variables, the number of lags and the number of the observations in the sample size. In these exercises in the real world, the same model used by Del Negro and Schorfheide (2004) is considered, hence the number of the endogenous variables do not change. The attention is focused on the sample size and obviously, as shown in the experiments in the artificial world, on the lag length.

The first exercise is a replication of Del Negro and Schorfheide (2004) with the same features, 80 quarters observations, the same number of lags, but what is changed is the lambda grid. This new set of lambda is more precise around the possible optimal lambda (in Del Negro and Schorfheide exercise and in this replication, $\widehat{\lambda} = 0.6$) $\Lambda = \{0, 0.2, 0.4, 0.5, 0.6, 0.7, 1, 1.4, 1.8, 10\}$[20]. In the second exercise, the time span is extended to include earlier years from the first quarter of 1961 and ends with the last quarter of 2001; i.e. the new sample size considers 160 observations. It is obvious that the lambda grid should change. The minimum λ is 0.1, instead of 0.2. The new lambda grid is $\Lambda = \{0, 0.10, 0.15, 0.2, 0.4, 0.5, 0.6, 0.7, 1, 1.4, 1.8, 10\}$[21]. In this exercise, the optimal λ is

[20] 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 it is not interesting the analysis of the shape of the marginal likelihood function.

[21] In this case, the infinite lambda representing the case in which restrictions from DSGE model are

0.3, instead of 0.6.

The following tables, the results of these two experiments are presented:

TABLE 12. VAR(4) Sample 1981-2001[22]		
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

extremely tight is not considered since in this experiment it is not interesting the analysis of the shape of the marginal likelihood function.

[22] 25,000 replications in the Metropolis-Hastings are implemented.

TABLE 13. VAR(4) Sample 1961-2001[23]		
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

Beyond taking into consideration the minimum and the optimal λ, it could be useful to consider the improvement of the theoretical model over the statistical model, the ratio $\frac{\widehat{\lambda}-\lambda_{MIN}}{\lambda_{MIN}}$.

In case of the small and the large sample, the ratio is equal to 2. Consequently, a larger sample size influences only the minimum and the optimal λ, but not the quality of explanation of the economic model.

Referring to MonteCarlo experiments, it is interesting to see what happens on the real data when there is an increase of the lags in VAR representation of the combination

[23] 25,000 replications in the Metropolis-Hastings are implemented.

DSGE-VAR.

According to Del Negro et al. (2007b), the lag length could be chosen by maximizing the marginal data density associated with the DSGE-VAR ($\widehat{\lambda}$). In this way, it is possible to consider not only VAR representation, but also the information from the economic model given by priors and the cross-moments. In the next table, there is the description of the $\widehat{\lambda}$[24] and its marginal data density, considering lags from 1 to 8.

TABLE 14. 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.

It is possible to consider the usual information criteria[25], the three real time series of U.S. economy could be represented in a VAR framework by considering a more parsimonious model. Schwarz and Hannan-Quinn criteria suggest the use of only one lag,

[24] 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 in the Metropolis-Hastings are implemented.

[25] 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.

while Likelihood Ratio, Final Prediction Error and Akaike criteria suggest the necessity of six lags. Consequently, both maximizing the marginal data density and using the usual information criteria on the real data suggest a more parsimonious representation[26].

Following these results, in the third exercise the most parsimonious VAR representation with only one lag is considered.

The next table shows the marginal data density of a VAR representation with only one lag in case of a small sample of 80 quarters and in case of a large sample of 160 quarters.

[26] It is true that considering AIC and SIC on the real data, the econometrician forgets 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 the econometrician 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.

TABLE 15. VAR(1) Two Samples[27]			
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

It is evident that changing the number of lags, the marginal likelihood changes its shape[28].

In the following table, the ratios calculated in case of VAR with four lags are compared

[27] 25,000 replications in the Metropolis-Hastings are implemented.

[28] 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.

with the ratios computed in case of a more parsimonious VAR.

TABLE 16. 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

Looking at the Table 16, it is clear that in case of VAR representation with only one lag, the ratio representing the improvement of DSGE over VAR, is smaller in the small sample size than in the large sample.

By using a more parsimonious statistical model, the explanation of the data by the economic model becomes less important.

Hence, it is crucial to understand under which criteria a VARMA representation of an economic model is truncated by using a VAR representation.

3.4.1 Forecasting

As done with MonteCarlo experiments, this exercise is completed by a forecasting evaluation.

The next tables analyze the forecasting performance for different sample size. As before, the small sample size considers 80 quarters from 1981:01 to 2001:04. Instead, the large sample considers 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 17. 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 18. 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 in the previous tables, 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 the next tables, 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 19. 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 20. 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.

4 Concluding Remarks and comments

This paper reconfirms that the interesting econometric tool, DSGE-VAR, developed by Del Negro and Schorfheide (2004) is very useful to understand if an economic model can explain real data.

MonteCarlo experiments with artificial data stress out the opportunity to set the lag length, since adding more lags than the Data Generating Process lag length could suggest that the economic model is better than the statistical framework representing the data. This result could be admissible if the data come from the same forward-looking model used as the candidate model in the hybrid composition DSGE-VAR. But this result is not so obvious if the data come from an alternative model, such as a backward-looking model.

The crucial point is how the VARMA representation for the economic model is truncated.

However, in the forecasting exercises, in case of using the forward-looking model, the best model is the DSGE-VAR with 3 lags and in case of using the backward-looking model, there is no a clear evidence on the best model when different lags are considered.

The exercises realized in the real world in order to assess the results provided by Del Negro and Schorfheide procedure,changing the sample size and the lag length, are useful to understand a crucial point. This point is the choose of the right number of lags in VAR representation. Moreover, this paper provides the importance of the use of DSGE-VAR in the real data, considering the forecasting performance.

Following the provided results, it is important to be careful to truncate the VARMA representation for the theoretical model, when not all the endogenous variables are observed, considering the more parsimonious representation. At this point, it is possible to maximize the marginal data density on the overall DSGE-VAR or applying the classical information criteria on the real data.

The next steps in the researcher agenda on this argument should be focused on a deeper analysis of the marginal likelihood. From the properties of the marginal likelihood it is possible to understand what effectively happens when more lags are added. Moreover, it could be necessary to analyze in depth the possible relationship between the usual information criteria and the ratio composed by the optimal and minimum λ.

In this way, it could be possible the use of a specific information criteria which depends on the priors of the model too, since the issue of the approximation of VARMA representation of a DSGE model is an important aspect to understand how DSGE-VAR approach works.

It is obvious that the hybrid model, DSGE-VAR, should be considered in its application under this new point of view, moreover, for using it in model validation in the DSGE approach.

References

[1] Adjemian, Stéphane, Matthieu Darracq Pariès and Stéphane Moyen (2008): "Towards a Monetary Policy Evaluation Framework", *ECB WP Series, No 942*.

[2] 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.

[3] An, Sungbae, and Frank Schorfheide (2007): "Bayesian Analysis of DSGE Models", *Econometric Reviews,* 26 (2-3): pp 1-60.

[4] Chari, V.V., Patrick Kehoe and Ellen R. McGrattan (2007): "Are Structural VARs with Long-Run Restrictions Useful in Developing Business Cycle Theory?", *Federal Reserve Bank of Minneapolis, Research Department Staff Report, 364*.

[5] Christiano, Lawrence J. (2007): "Comment on Marco Del Negro, Frank Schorfheide, Frank Smets and Raf Wouters "On the Fit of New-Keynesian Models"", *Journal of Business and Economic Statistics*, Volume 25, Number 2 April 2007, pp.143-151.

[6] 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.

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

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

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

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

[11] 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 .

[12] 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 .

[13] Faust, Jon (2007): "Comment on Marco Del Negro, Frank Schorfheide, Frank Smets and Raf Wouters "On the Fit of New-Keynesian Models"", *Journal of Business and Economic Statistics*, Volume 25, Number 2 April 2007, pp154-156.

[14] 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.

[15] Gallant, A. Ronald (2007): "Comment on Marco Del Negro, Frank Schorfheide, Frank Smets and Raf Wouters "On the Fit of New-Keynesian Models"", *Journal of Business and Economic Statistics*, Volume 25, Number 2 April 2007, pp151-152.

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

[17] Hansen, Lars Peter (2004): "An Interview with Christopher A. Sims", *Macroeconomic Dynamics*, 8, pp 273-294.

[18] 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.

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

[20] Kass, Robert E. and Adrien E. Raftery (1995): "Bayes Factors", *Journal of the American Statistical Association*, Vol. 90, No. 430, pp. 773-795.

[21] Kilian, Lutz (2007): "Comment on Marco Del Negro, Frank Schorfheide, Frank Smets and Raf Wouters "On the Fit of New-Keynesian Models"", *Journal of Business and Economic Statistics,* Volume 25, Number 2 April 2007, pp 156-159.

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

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

[24] Laforte, Jean-Phillipe (2003): "Comparing Monetary Policy Rules in an Estimated General Equilibrium Model of the U.S. Economy", Manuscript, Federal Reserve Board of Governors, Washington, DC.

[25] Lees, Kirdan, Troy Matheson and Christie Smith (2007): "Open Economy DSGE-VAR Forecasting and Policy Analysis: Head to Head with the RBNZ Published Forecasts", *Discussion Paper Series* DP2007/01 Reserve Bank of New Zealand.

[26] Levin, Andrew, Volker Wieland and John C. Williams (2003): "The Performance of Forecast-Based Monetary Policy Rules Under Model Uncertainty", *American Economic Review*, 93, 622-645.

[27] 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.

[28] Liu, Guangling Dave, Rangan Gupta and Eric Schaling (2008): "Forecasting the South African Economy: A DSGE-VAR Approach", *Center Tilburg University Discussion Paper* No. 2008-32.

[29] Lucas, Robert. E. Jr. (1976): "Econometric Policy Evaluation: A Critique", In K. Brunner and A. Meltzer (eds.) *The Phillips curve and labor markets.* Amsterdam: North-Holland.

[30] MacKay, David J.C. (2003): "Information Theory, Inference & Learning Algorithms", Cambridge University Press.

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

[32] Robert, Christian P. (2007): "The Bayesian Choice:From Decision-Theoretic Foundations to Computational Implementation", Second Edition, Springer.

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

[34] Schorfheide, Frank (2000): "Loss Function-Based evaluation of DSGE Models", *Journal of Applied Econometrics*, 15, S645-670.

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

[36] Schwarz, Gideon (1978): "Estimating the Dimension of a Model", *Annals of Statistics* 6(2): pp 461-464.

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

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

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

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

[41] Sims, Christopher A. (2007): "Comment on Marco Del Negro, Frank Schorfheide, Frank Smets and Raf Wouters "On the Fit of New-Keynesian Models"", *Journal of Business and Economic Statistics*, Volume 25, Number 2 April 2007, pp152-154.

[42] 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.

[43] Spiegelhalter, David J., Nicola Best. and Bradley P. Carlin (1998): "Bayesian Deviance, the Effective Number of Parameters, and the Comparison of Arbitrarily Complex Models", Research Report 98–009, Division of Biostatistics, University of Minnesota.

[44] Spiegelhalter, David J.; Nicola G. Best, Bradley P. Carlin, and Angelika Van Der Linde (October 2002): "Bayesian measures of model complexity and fit (with discussion)", *Journal of the Royal Statistical Society*, Series B (Statistical Methodology) 64 (4): pp 583–639.

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

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

5 Appendix

APPENDIX. Summary Table 1000 observations sample

	λ min	λ opt	λopt-λmin	$(\lambda opt - \lambda min) / \lambda$ min
DSGE-VAR(1)	0.007	0.01	0.003	0.43
DSGE-VAR(2)	0.01	0.016	0.006	0.6
DSGE-VAR(3)	0.013	0.02	0.007	0.54
DSGE-VAR(4)	0.016	0.02	0.004	0.25
DSGE-VAR(5)	0.019	0.028	0.009	0.47
DSGE-VAR(6)	0.022	0.03	0.008	0.36
DSGE-VAR(7)	0.025	0.03	0.005	0.2
DSGE-VAR(8)	0.028	0.03	0.032	1.14

APPENDIX. Maximizing Marginal Data Density in one replication of MC

	λ min	λ opt	MDD
DSGE-VAR(1),T=80	**0.09**	**0.13**	**-21.783**
DSGE-VAR(2), T=80	0.13	0.17	-31.564
DSGE-VAR(3), T=80	0.17	0.20	-39.829
DSGE-VAR(4), T=80	0.20	0.31	-44.515
DSGE-VAR(5), T=80	0.24	0.31	-50.125
DSGE-VAR(6), T=80	0.28	0.35	-54.301
DSGE-VAR(7), T=80	0.31	0.5	-56.431
DSGE-VAR(8), T=80	0.35	0.7	-58.111

APPENDIX. 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 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						