UNIVERSITÀ COMMERCIALE LUIGI BOCCONI

PHD IN ECONOMICS
CICLO XX

MODEL VALIDATION IN THE DSGE APPROACH

ALESSIA PACCAGNINI
Matricola 1003904

COMMISSION:

Prof. Carlo Ambrogio Favero, Università Commerciale Luigi Bocconi
Prof. Luca Sala, Università Commerciale Luigi Bocconi
Dr. Marco Del Negro, New York Federal Reserve

Contents

Preface

In these days, I have read my dissertation several times in order to be sure to have written all I would like to tell the reader, regarding the Model Validation in the DSGE Approach.

What were the reasons that drove me to write this thesis?

First of all, my huge passion for the Econometrics, a great love began at the undergraduate time! Second, my curiosity and desire to apply the Econometrics to the economics model, in order to understand if a model can explain the real data! Third, my interest in the what is the frontier of the research.

It is obvious that in the choice of the issue of my dissertation I was not lonely. My supervisors, the courses I have done, the papers I have read conducted me on this field. And every day, my determination, my curiosity and my passion for the research led me to write these papers, included in this thesis.

The possibility of applying the Bayesian Econometrics tools to the modern Macroeconomics models has been an interesting challenge. Moreover, I had the opportunity to study very recent methods and applications of what is the so-called "New Macroeconometrics".

In a recent paper of Jesus Fernandez-Villaverde (2009), he states: ". in the New Macroeconometrics.., we need more eager young minds to join us. I hope that some readers will find this call intriguing".

For my part, I hope to be among these young minds and I hope I can make a small contribution to research in this field, not only with this thesis, but with my future projects.

Acknowledgements

I am grateful to the invaluable support and guidance from my supervisors Prof. Carlo A. Favero and Prof. Luca Sala. I thank Dr. 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 discussions with Prof. Barbara Chizzolini, Prof. Marco Maffezzoli, Prof. Massimiliano Marcellino, Prof. Giorgio Primiceri, Prof. Bruno Sitzia and Prof. Ulf Sodestrom. I thank participants at Università Bocconi Ph.D. Student Workshops, ADDEGeM 2008, Doctoral Meeting of Montpellier (D.M.M.) 1st edition (2008), First Italian Ph.D. Student Workshop, Collegio Carlo Alberto, Turin (2008), EEA-ESEM Milan, Università Bocconi (2008) and RES Fourth PhD Presentation Meeting in London (2009).

I would like to thank the Director of the Ph.D. School, Prof. Pierpaolo Battigalli and the all Faculty for providing me excellent education. I thank the Italian Ministry of Education and the Department of Economics for the fellowships.

I am particularly indebted to the secretaries who helped me to overcome with the bureaucracy duties, in particular, I would like to thank Angela Baldassarre, Marina Catenacci, Alessandra Gadioli, Paola Pasinetti, Marika Podelli, Nicola Scalzo and Alessandra Startari.

I thank my undegraduate and Master friends for having supported over the years, Elisa Borghi, Rosario Crinò and especially Italo Colantone.

I would like to thank my Ph. D. colleagues, I shared with them problems, pains, especially laughter. In particular, I thank Elena Besedina, Paolo Bianchi, Michela Braga, Marianna Caccavaio, Agostino Consolo, Lucia Corno, Paolo Di Giannatale, Valerio Ercolani, Valeria Gattai, Linlin Niu, Sara Pinoli and Mario Porqueddu.

A special thank to Agostino, I was so glad to work with him in one of my papers!

I cannot forget the friends of the offices 411 and 5 e3-03, Claudia Foroni, Marcella Nicolini and Riccardo M. Masolo (and obviously again Sara, Mario and Paolo Bianchi!). I thank them for their warm support and for the beautiful moments spent together. I had I hope that our friendship can develop into a collaboration for our future papers!

Another special thank to Marcella, not only to for sharing the office, but also to be my coauthor in one of my papers which combines our passions for econometrics and international economics (paper not included in this dissertation!).

I would like to thank my students, in particular students of Advanced Econometrics course, they were my first audience when I tried to explain in "easy" words my research projects!

Last but not least, huge thanks to my family: my mother, my father and my grandmothers to support me and encourage me to follow my dreams!

I remain solely responsible for all errors, omissions, and interpretations.

Introduction

Over the last few years, Dynamic Stochastic General Equilibrium Models (DSGE) are considered very attractive in Academia, Central Banks and international policy institutions, not only for theoretical exercises, but also for the empirical applications to explain macroeconomic fluctuations and conduct quantitative policy analysis. The estimation and the evaluation of DSGE models are considered as two crucial aspects in the mainstream New Macroeconometric literature.

This thesis focuses on the model validation in the DSGE approach, considering the problems of the statistical representation and of the approximation of these theoretical models by using recent econometrics tools. The dissertation is composed of three chapters, which are represented by three papers, readable independently.

The first chapter entitled as "Model Validation in the DSGE Approach: What is Done and What is Next?" is a critical survey of the recent literature on the model validation in the DSGE approach. This paper concentrates the attention on the "New Macroeconometrics", especially, on the application of the Vector Autoregressions (VARs) as a representation for the theoretical model. A special emphasis is given on the use of the Bayesian methods, not only for obtaining the posterior distributions for the parameters of the model, but also for the promising application in hybrid or mixture models. The introduction of the dummy observation priors in the macroeconomic research has allowed the economist to create models which are a combination of VAR representation for the actual data and Bayesian VAR representation for the artificial data (which are the so-called "dummy observation priors"). One promising example of these mixture models is the DSGE-VAR proposed by Del Negro and Schorfheide (2004). In the first chapter, the problems of the statistical representation of a theoretical model (Spanos

(1990)) and of the approximation of a DSGE model by using VARMA or finite-order VAR representation (Fernandez-Villaverde et al. (2007) and Ravenna (2007) are discussed by using two different models. A particular attention is given to the recent application of DSGE-VAR, considering these crucial points. In this sense, it is possible to introduce the contribution of the two other chapters of this thesis. The contribution of the second chapter regards the statistical representation of DSGE-VAR compared to a proposed DSGE-FAVAR. Instead, the contribution of the third chapter concerns the problem of the approximation of VARMA representation in an hybrid model The critical survey concludes proposing a new application of hybrid model, considering the importance of the structural component of DSGE models.

The second chapter, is entitled as "On the Statistical Identification of DSGE Models[1]". This paper is a joint work with Agostino Consolo and Carlo A. Favero and it has been revised and resubmitted to the Journal of Econometrics.

This chapter evidences as a consequence of the development of the application of DGSE, methods for diagnosing the fit of these models are being proposed and implemented. In this article we illustrate how the concept of statistical identification, that was introduced and used by Spanos (1990) to criticize traditional evaluation methods of Cowles Commission models, could be relevant for DSGE models. We conclude that the recently proposed model evaluation method, based on the DSGE-VAR, might not satisfy the condition for statistical identification. However, our application also shows that the adoption of a FAVAR as a statistically identified benchmark leaves unaltered the support of the data for the DSGE model and that a DSGE-FAVAR can be an optimal forecasting model.

The last chapter, but not least, entitled as "Assessing the Potential of DSGE Model Evaluation in a Bayesian Framework[2]", highlights the problems of the use of hybrid model, such as the DSGE-VAR. This paper proposes two different empirical contributions.

The first empirical analysis is realized by using MonteCarlo experiments in the artificial world. The purpose of the first MonteCarlo experiment is to evidence if the DSGE-VAR approach is useful to understand if the artificial data generated by the theoretical model can be

[1] I presented this paper at the EEA-ESEM 2008 Conference in Milan. Instead, my coauthors presented this paper at CREI Conference 2007, in Barcelona and DIW Macroeconometric 2008 Workshop in Berlin.

[2] This paper has been presented at the ADDEGeM 2008, Doctoral Meeting of Montpellier (D.M.M.) 1st edition; at First Italian PhD Student Workshop 2008, Collegio Carlo Alberto, Turin and at the Fourth PhD Presentation Meeting 2009, in London.

explained by this model. In this sense, the data generating process is the forward-looking model used in the combination for generating the dummy observation priors. This exercise shows how the DSGE-VAR is able to recognize that the DGP is the theoretical model and misspecifying the approximation of the VARMA representation of the model, it seems there is more contribution of the DSGE to explain the artificial data. The second MonteCarlo experiment is to evidence if the DSGE-VAR approach is useful to recognize if the data generating process is not the model used to generate the dummy observation priors. In this exercise, the DGP is a backward-looking model which is a different model respect the model used in the combination. The results of this experiment are interesting, it can happen that the DSGE-VAR recognizes that DGP is the model used to generate the dummy observations, misspecifying the lags approximation of the VARMA representation of the model. In this sense, it is very important to be careful in the choice of the finite-order VAR representation.

The second empirical analysis regards the real world. In this exercise, the application of DSGE-VAR on real U.S. data is realized, following Del Negro and Schorfheide (2004) and following the results of MonteCarlo experiments. It is possible to evidence that the use of a more parsimonious VAR representation suggests less contribution for the economic model and in this sense it is important to choose the correct lag-length

Considering these results, it is necessary to apply the DSGE-VAR considering the same comments given to the use of VAR representation. Moreover, the problem of the misspecification of the number of lags in the econometric representation causes an increase in the marginal likelihood function in the procedure of the combination, hence the hybrid model is not able to recognize this problem. In this point of view, it is necessary the use of Information Criteria on the actual data or the maximization of the marginal data density on the overall DSGE-VAR, in order to check the correct approximation. This problem can be also considered as a starting point for a further analysis of the approximation of theoretical models, using a correct methodology.

Chapter 1

Model Validation in the DSGE Approach: What is Done and What is Next?

Abstract

In this paper a review of the recent literature on the Model Validation in the Dynamic Stochastic General Equilibrium (DSGE) Models approach is discussed. The paper is focused on the "New Macroeconometrics" and on the application of VAR representation for the theoretical model. A special emphasis is given on the application of the Bayesian methods for the combination of hybrid models, such as the DSGE-VAR. The paper concludes the review considering some pointers for the future research and for the further developments of the use of the dummy observation priors in the macroeconomic empirical analysis.

Keywords: Bayesian Analysis, DSGE Models, Vector Autoregressions

JEL Classification: C11, C15, C32

"Dynamic equilibrium theory made a quantum leap between the early 1970s and the late 1990s. In the comparatively brief space of 30 years, macroeconomists went from writing prototype models of rational expectations (think of Lucas, 1972) to handling complex constructions like the economy in Christiano, Eichenbaum, and Evans (2005). It was similar to jumping from the Wright brothers to an Airbus 380 in one generation".

Jesus Fernandez-Villaverde in "The Econometrics of DSGE Models"(2009)

1.1 Introduction

Dynamic Stochastic General Equilibrium Models (DSGE) are becoming very used in academics, international policy institutions and Central Banks research projects in order to explain the macroeconomic fluctuations and conduct quantitative policy analysis. These models are based on new-Keynesian framework and the term DSGE was originally used by Kyndland and Prescott (1982) in their seminal contribution on Real Business Cycle (RBC) model. The RBC model is based on neoclassical framework with micro-founded optimization behavior of economic agents with flexible prices. DSGE models have the advantage to combine micro-foundations of both households and firms optimization problems and with a large collection of both nominal and real (price/wage) rigidities. These models allow to establish a link between structural features of the economy and reduced form parameters, something that was not always possible with the usual large-scale macroeconomic models. However, it is possible to argue that the Lucas Critique (1976)[1] was influential to encourage macroeconomists to build microfoundations for their models.

The DSGE models have been very attractive in policy analysis, thanks to the use of the calibration methods. In the last years, the popularity of the calibration has declined since the improvements in computational power and the development of new econometric methods. It is obvious that the combination of rich structural models, novel solution algorithms and

[1]Lucas (1976) and Lucas and Sargent (1979) suggest that if an economist want to predict the effect of a policy experiment, she should model the "deep parameters" (relating to preferences, technology and resource constraints, the parameters which do not vary with policy) that govern individual behavior. It is possible to predict what individuals will do taking into account the change in policy, and then aggregate the individual decisions to calculate the macroeconomic effects of the policy change.

powerful simulation techniques has allowed researchers to develop what is the so-called "New Macroeconometrics" (Fernandez-Villaverde (2009)). As discussed by Schorfheide (2008), the term "macroeconometrics" is often narrowly associated with large-scale system-of-equations models in the Cowles Commission tradition which were developed from 1950s to the 1970s. Among the most important attacks to these models, there were the famous Lucas critique (1976) and Sims (1980) seminal paper introducing the Vector Autoregression (VAR) approach. Sims (1980) criticized that many restrictions that are used to identify behavioral equations in these models are inconsistent with dynamic macroeconomic theories and proposed the use of the VARs.

There are several contributions of different econometric techniques available for estimating DSGE models. Examples of these include estimation of equilibrium relationships with generalized method of moments (GMM), application of Structural VARs, minimum distance estimation based on the discrepancy between VAR and DSGE impulse response functions, maximum likelihood and Bayesian Econometrics (see Canova (2007), An and Schorfheide (2007), Ruge-Murcia (2007) for detailed reviews of the different approaches).

However, what is very important in this field is the model validation issue. The most important contribution in the use of the model validation is represented by Spanos (1990). The concept of the model validation refers to the possibility to assess the statistical identification of a model, in this case of a DSGE model. There are two main steps in this analysis, the first one is the statistical representation of the model and second, the use of this representation in order to explain the data.

The concept of statistical identification has been introduced by Spanos (1990). Structural models can be viewed statistically as a reparameterization, possibly (in case of over-identified models) with restrictions, of the reduced form. Spanos distinguishes between structural identification and statistical identification. Structural identification refers to the uniqueness of the structural parameters, as defined by the reparameterization and restriction mapping from the statistical parameters in the reduced form, while statistical identification refers to the choice of a well-defined statistical model as reduced form. Diagnostics for model evaluation are constructed in Cowles commission tradition in a way that is closely related to the solution of the identification problem. In fact, in the (very common) case of over-identified models, a test of the validity of the over-identifying restrictions can be constructed by comparing the restricted reduced form implied by the structural model with the reduced form implied by the just-identified model in which each endogenous variables depend on all exogenous variables with unrestricted coefficients. The statistics are derived in Anderson and Rubin (1949) and Basman (1960). The logic of the test attributes a central role to the structural model. The statistical model of reference for the evaluation of the structural model is derived by the structural model itself. Spanos (1990) points out that the root of the failure of the Cowles Commission approach lies in the little attention paid to the statistical model implicit in the estimated structure. Any identified structure that is estimated without checking that the implied statistical model is an accurate description of the data is bound to fail if the statistical model is not valid. The Spanos critique of the Cowles commission approach lies naturally within the LSE approach to econometric modelling. Such approach reverses the prominence of the structural model with respect to the reduced form representation. The LSE approach starts its specification and identification procedure with a general dynamic reduced form model. The congruency of such a model cannot be directly assessed against the true DGP, which is unobservable.

The basic and the most used econometric tool for the empirical validation of macroeconomic models is the Vector Autoregressive Model (VAR). Traditionally, this model is easy to estimate and, once identification restrictions are imposed, it can be used to evaluate the impact of economic shocks on key variables. Instead, in structural VAR macroeconomics, variables are represented as driven by serially uncorrelated shocks, each having a different source or nature, like “demand”, “supply”, “technology”, “monetary policy” and so on. Each variable reacts to a particular shock with a specific sign, intensity and lag structure, summarized by the so-called “impulse-response function”. Implications of economic theory not used for identification can then be compared with estimation results and tested. The application of the SVAR approach in the empirical macroeconomics has been growing in order to uncover economic relationships without imposing strong theoretical assumptions as in the use of the reduced VAR. SVARs have been attractive for macroeconomist to emphasize on the structural shocks embodied in the complex and modern DSGE models.

However, there are several papers debating the failure of the applications of this econometric tool. A strong motivation for the use of VARs is that DSGE models have solutions that can be represented in VAR form and therefore VAR econometrics provide the tool to bridge theory and data. One another important point in this discussion is the problem of the variables which are presented in a general forward-looking model, but they are not observed. For example, the autoregressive government spending or technology shocks which do not correspond to the observable variables.

At the same time, the criticisms against the implementation of SVARs proposed the use of the reduced form in order to approximate the DSGE model with unobservable variable, which can be represented as a VARMA. In this sense, the literature has discussed the use and the conditions to choose the best VAR representation for a theoretical model. Moreover, these papers have discussed the assumptions needed for a finite-order VAR representation of a DSGE model to exist (Fernandez-Villaverde, Rubio-Ramirez, Sargent and Watson (2007) and Ravenna (2007)).

When a VAR(p) is only an approximation to the true VAR, the truncated VAR(p) may return largely incorrect estimates of the impulse responses, of the forecasting evaluation. This problem is crucial both in estimation exercises and in policy simulation. But what is the big improvement in the recent macroeconometric field, as said before, called "New Macroeconometrics", is the use of the Bayesian Econometrics. This literature on Bayesian estimation of DSGE models began with works by Landon-Lane (1998), DeJong, Ingram and Whiteman (2000), Schorfheide (2000) and Otrok (2001), followed by Smets and Wouters (2003, 2007).

These new econometrics tools applied to the macroeconomics allow the economist to consider the model validation in the DSGE approach.

The use of the Bayes' theorem is not useful only to combine priors of the parameters of a DSGE model with the likelihood, obtaining the posterior distributions of these parameters. One of the advantages of the application of Bayesian methods is the opportunity to connect behavioral component of the theoretical model, such as macroeconomic models (Real Business Cycle or Dynamic Stochastic General Equilibrium), with statistical models that fit well the data (Ingram and Whiteman (1994); DeJong et al.(1996, 2000), Del Negro and Schorfheide (2004); Del Negro, Schorfheide, Smets and Wouters (2007), Del Negro, Diebold and Schorfheide (2008) and Sims (2008)). Hence, the Bayesian econometrics helps the economist to estimate the real data by using mixture models which are a combination between the real data and the artificial data (the so-called "dummy observation priors") which are derived by the restrictions implied on the theoretical model. The application of this kind of mixture model is useful to implement the model validation in the DSGE approach

In the use of these hybrid models, there are two important points to consider. First, the concept of the statistical identification (Spanos, 1990) could be applied to the diagnostic tools proposed for DSGE models (Consolo, Favero and Paccagnini, 2007). Second, it needs to be careful in the approximation of the DSGE model by using VAR representations. It seems that mispecifying the number of lags, there is more weight for the theoretical model in the combination (Paccagnini, 2009).

Several extensions of the use of the mixture models are proposed. The more promising application of hybrid model is the DSGE-VAR introduced by Del Negro and Schorfheide (2004). According to Sims (2008), in this procedure one of the missing points is the analysis of the long-run structural component of the DSGE model. Moreover, in Del Negro et al. (2008), the use of the DSGE-VAR combination is associated with the frequency analysis, in order to distinguish the low and high frequencies.

The contribution of this paper is to analyze what are the critical aspects in the model validation in the DSGE approach and proposing some pointers for the future research and discuss eventual solutions following the examples of the recent contributions in the macroeconometric literature.

The remainder of the paper is organized as follows. In Section 2, the mixture models used for the model validation in the DSGE approach are presented. In Section 3, the problem of the statistical identification (following Spanos (1990)) is discussed by using a simple small-scale New-Keynesian model with all observable variables. In Section 4, the problem of the approximation of DGSE models (considering Fernandez-Villaverde et al. (2007) and Ravenna (2007)) is presented by using a simple small-scale New Keynesian model with some variables which are not observable. In Section 5, the future research agenda is presented. Concluding remarks are in Section 6.

1.2 Mixture Models

An important point of the use of a statistical representation for the data, such as Vector Autoregressive model (VAR), is the "overfitting" due to the inclusion too many lags and too many variables, some of which may be insignificant. The problem of "overfitting" results in multicollinearity and loss of degrees of freedom, leads to inefficient estimates and large out-of-sample forecasting errors. It is possible to overcome this problem by using Bayesian techniques. The main motivation for the use of this methodology is based on the knowledge that more recent values of a variable are more likely to contain useful information about its future movements than older values. This point is realized by introducing the use of dummy observation priors. The basic idea of dummy observation priors for VAR or in general for regression models is the addiction of "extra data" to the sample of the real data in order to express prior beliefs about the parameters. This curse of dimensionality is overcome by the shrinkage of the parameter space and imposing the prior beliefs on the parameters. The parameter space can be shrunk by imposing a set of restrictions, which could be for instance obtained from a theoretical structural model, directly on the parameters. Alternatively, one could use techniques, where prior distributions are imposed on the parameters to be estimated.

The first examples of this application (Goldberg and Theil, 1961; Litterman, 1981; Doan et al., 1984 and Litterman, 1986) are a Bayesian model that imposes restrictions on those coefficients by assuming they are to be near zero, by using normal prior.

Bayesian VARs (BVARs) proposed by Doan, Litterman and Sims (1984), for example, use what has become know as "Minnesota" priors to shrink the parameters space. The basic principle behind this procedure is that all equations are centered around a random walk with drift. This idea has been modified in Kadiyala and Karlsson (1997) and Sims and Zha (1998).

In Ingram and Whiteman (1994), a RBC model is used to generate a prior for a reduced form VAR, developing the "Minnesota" priors procedure. The key element in Ingram and Whiteman's analysis is that the dimension of the observable vector exceeds the dimension of the state vector. This representation is compared to a structural VAR (SVAR) representation and it holds under the condition that the number of the observable variables exceed those in the state vector, can provide information only for the first lag of a VAR(p). At this point, it is used an assumption that the rest (p-1) blocks of autoregressive parameters are normally distributed with mean zero and a covariance matrix that it is proportional to the covariance matrix of the first lag with the proportion parameter being an inverse function of the number of lags.

In DeJong et al. (1996, 2000), a prior is placed on parameters of a simple linearized DSGE, which is then compared with a BVAR in a forecasting exercise. This approach has advantages providing the econometrician with more complete stories about what behavioral mechanisms produce a given forecast, forecast error or policy scenario. There are also some disadvantages, it seems that these models more often than not fail to fit as well as models with little or no behavioral structure. Moreover, Smets and Wouters (2003) extend this last approach to medium scale New Keynesian models used in Central Bank policy analysis.

In Del Negro and Schorfheide (2004) and Del Negro et al. (2007), a DSGE prior is developed for a VAR, like Ingram and Whiteman (1994). They proposed a way to use DSGE model in order to generate a prior distribution for a structural time series model that relaxes the tight theoretical restrictions of the DSGE.

In this sense, the theoretical model is treated as a mechanism for generating prior distributions for the parameters in the unrestricted VAR. 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. (2007), this λ represents 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.

The DSGE-VAR hybrid model, proposed by Del Negro and Schorfheide (2004), is based on the rational expectations solution of the linearized model, computed using the algorithm implemented by Sims (2002).

As discussed in the previous section, the DSGE model can be represented by using the state-space form solution. Adopting the notation in Fernandez-Villaverde et al. (2007):

$$x_{t+1} = A(\theta)x_t + B(\theta)\varepsilon_t \quad (1.1)$$
$$y_t = C(\theta)x_t + D(\theta)\varepsilon_t$$

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 shock s is equal to the number of observable variables.

In DSGE-VAR, one of the most important aspect is the finite-order VAR approximation to the DSGE model. 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:

$$y_t = \sum_{j=1}^{\infty} C\left(A - BD^{-1}C\right)^{j-1} BD^{-1} y_{t-j} + D\varepsilon_t \quad (1.2)$$

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:

$$\Phi^*(\theta) = \Gamma^{*-1}_{xx}(\theta)\,\Gamma^*_{xy}(\theta) \quad (1.3)$$
$$\Sigma^*(\theta) = \Gamma^*_{yy}(\theta) - \Gamma^*_{yx}(\theta)\,\Gamma^{*-1}_{xx}(\theta)\,\Gamma^*_{xy}(\theta)$$

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:

$$Y^* = X^*\Phi(\theta) + U^* \tag{1.4}$$

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.

Following the Bayesian Econometrics literature that suggests the use of dummy observations to impose a prior distribution on the set of coefficients, in the Del Negro and Schorfheide procedure, these dummy observation priors 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.

The DSGE-VAR has not restrictions imposed "dogmatically" as in Ingram and Whiteman (1994). 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 the vector the covariance matrix Σ_u. This 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.

1.3 A Simple DSGE Model: the Statistical Representation

As discussed in the Introduction, the use of an unrestricted VAR is one of the most promising way to represent the DSGE as a statistical model. If the unrestricted VAR is not statistically identified the econometrician might not be able to conduct a proper statistical evaluation of the theoretical model by using it as a benchmark. The important point made by Spanos (1990) in his contribution is that reduced form of theoretical must be evaluated against specifications that captures the relevant information in the data potentially omitted from the theoretical model. Actually, there are a number of potential sources of mis-specification for the model derived VAR. Think of all those variables that are omitted from the theoretical model because of its specific nature, say fiscal policy in a model designed to analyze the effect of monetary policy or foreign variables and the exchange rate in a closed economy model, but also of all variables that are not theory related but are relevant to determine the actual behavior of policy makers. A good example is the Commodity Price Index (CPI) and the problem of modelling of the behavior of monetary policy authority. Early VARs for the analysis of monetary policy that did not include in the information set a commodity price index tended to deliver a "price puzzle", i.e. a positive response of prices to an unexpected monetary tightening. Such anomaly has been attributed to the existence of a leading indicator for inflation to which the Fed reacts and which is omitted from the VAR. The omission form the information set of a variable positively correlated with inflation and interest rates makes the VAR mis-specified and explains the positive relation between prices and interest rates observed in the impulse response functions. It has been observed (see Christiano, Eichenbaum and Evans (1998)) that the inclusion of a Commodity Price Index in the VAR solves the 'price puzzle'. DSGE model do not typically include the commodity price index in their specification as a consequence the VAR derived by relaxing the theoretical restrictions in a DSGE model is misspecified. So the evaluation of the effects of conducting model misspecification with a "wrong" benchmark is a practically relevant one.

For example, it possible to consider an example of a small-scale New-Keynesian model which suffers from the omission of the CPI. In this session a model with all observable variables is considered.

This 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 input 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, the government spending and the technology shock. These shocks are iid normal idiosyncratic 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:

$$\widetilde{x}_t = E_t[\widetilde{x_{t+1}}] - \frac{1}{\tau}(\widetilde{R_t} - E_t[\widetilde{\pi_{t+1}}]) + \epsilon_{g,t} \quad (1.5)$$

$$\widetilde{\pi}_t = \beta E_t[\widetilde{\pi_{t+1}}] + \epsilon_{z,t} \quad (1.6)$$

$$\widetilde{R_t} = \rho_R \widetilde{R_{t-1}} + (1-\rho_R)(\psi_1 \widetilde{\pi}_t + \psi_2 \widetilde{x}_t) + \epsilon_{R,t} \quad (1.7)$$

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, $\epsilon_{g,t}$.

The second equation represents the inflation dynamics determined by the expectional Phillips curve. 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 technology shock is given by $\epsilon_{z,t}$.

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.

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 :

$$\Gamma_0 \mathbf{Z}_t = \Gamma_1 \mathbf{Z}_{t-1} + C + \Psi \epsilon_t + \Pi \eta_t \tag{1.8}$$

$t = 1, ..., T$ 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 not given exogenously but treated as part of the model solution, satisfying $E_t \left(\eta_{t+1} \right) = 0$, all t.

In this model the result obtained by using the Sims' algorithm (2002) is:

$$\mathbf{Z}_t = \begin{bmatrix} \widetilde{x}_t \\ \widetilde{\pi}_t \\ \widetilde{R}_t \\ \widetilde{R}^*_t \\ E_t \widetilde{x_{t+1}} \\ E_t \widetilde{\pi_{t+1}} \end{bmatrix} \quad \epsilon_t = \begin{bmatrix} \epsilon^R_t \\ \epsilon^G_t \\ \epsilon^Z_t \end{bmatrix} \qquad \eta_t = \begin{bmatrix} \eta^x_t = x_t - E_{t-1}(x_t) \\ \eta^\pi_t = \pi_t - E_{t-1}(\pi_t) \end{bmatrix}$$

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

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

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

The forcing processes here are the elements of the vector ϵ_t, this typically contains processes like Commodity Price Index that is not determined by an optimization process. Policy variables set by optimization, typically included $\mathbf{Z}_t$, are naturally endogenous as optimal policy requires some response to current and expected developments of the economy. Expectations at time t for some of the variables of the systems at time t+1 are also included in the vector $\mathbf{Z}_t$,whenever the model is forward looking. Model like (2.1)can be solved using standard numerical techniques (see, for example, Sims, 2002), and the solution can be expressed as:

$$\mathbf{Z}_t = \mathbf{A}_0 + \mathbf{A}_1 \mathbf{Z}_{\mathbf{t-1}} + \mathbf{R}\epsilon_t \tag{1.9}$$

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

$$\mathbf{Z}_t = \mathbf{A}_0 + \mathbf{A}_1 \mathbf{Z}_{\mathbf{t-1}} + \mathbf{u}_t \tag{1.10}$$

Within this framework a new role for empirical analysis based on reduced form models emerges, that is to provide evidence on the stylized facts to be matched by the theoretical model adopted for policy analysis and to decide between competing DSGE models.

The identification of the shocks of interest is the structural identification problem in VAR-based model evaluation. VAR modelling recognizes that identification and estimation of structural parameters is impossible without explicitly modelling expectations, therefore a structure like (2.3) can only be used to run special experiments that do not involve simulating different scenarios for the parameters of interests. A natural way to achieve these results is to experiment with the shocks ϵ_t. *Facts* are then provided by looking at impulse response analysis, variance decompositions and historical decompositions. All these experiments are run by keeping estimated parameters unaltered. Importantly, running these experiments is easier if shocks to the different variables included in the VAR are orthogonal to each other, otherwise it would not be possible to simulate a policy shock by maintaining all the other shocks at zero. As a consequence, VAR models need a structure because orthogonal shocks are normally not a feature of the statistical model. This fact generates the structural identification problem. The relation between (2.3) and (2.2)implies that:

$$\mathbf{u}_t = \mathbf{R}\epsilon_t,$$

from which we can derive the relation between the variance-covariance matrices of $\mathbf{u}_t$ (observed) and $\boldsymbol{\nu}_t$ (unobserved) as follows:

$$E\left(\mathbf{u}_t \mathbf{u}_t'\right) = \mathbf{R} E\left(\epsilon_t {\epsilon_t}'\right) \mathbf{R}'. \tag{1.11}$$

Substituting population moments with sample moments the relevant structural shocks can be identified and a number of parameters in the $\mathbf{R}$ matrix equal up to the number of different elements in the variance covariance matrix of the VAR innovations ($m(m+1)/2$, where m is the number of variables included in the VAR) can be estimated. As usual, for such a condition also to be sufficient for identification, no equation in (2.4) should be a linear combination of the other equations in the system (see Amisano and Giannini 1996, Hamilton 1994). As for traditional models, we have the three possible cases of under-identification, just-identification and over-identification. The validity of over-identifying restrictions can be tested via a statistic distributed as a χ^2 with the number of degrees of freedom equal to the number of over-identifying restrictions. But again structural identification of a VAR is a totally different from statistical identification. In fact, misspecification of a VAR generated by omitted variables does not prevent structural identification but it leads to lack of statistical identification.

Consequently, the problem is that the DGP includes the Consumer Price Index, instead the VAR representation derived from the theoretical model omits this variable. A way to recover this variable is to represent the model by using a Factor-Augmenting VAR which nests the VAR representation.

Moreover, the use of Factor Augmented VAR deals with the so-called "curse of dimensionality" (Chudick and Pesaran, 2007), following the approach of the shrinkage of the data as introduced by Geweke (1977) and Sargent and Sims (1977) and applied by Stock and Watson (1999, 2002, 2005), Giannoni, Reichlin and Sala (2005), Bernanke, Boivin and Eliasz (2005) and Boivin and Giannoni (2006).

A FAVAR benchmark for the evaluation of the previous DSGE model will take the following specification:

$$\begin{pmatrix} \mathbf{Y}_t \\ \mathbf{F}_t \end{pmatrix} = \begin{bmatrix} \mathbf{\Phi}_{11}(L) & \mathbf{\Phi}_{12}(L) \\ \mathbf{\Phi}_{21}(L) & \mathbf{\Phi}_{22}(L) \end{bmatrix} \begin{pmatrix} \mathbf{Y}_{t-1} \\ \mathbf{F}_{t-1} \end{pmatrix} + \begin{pmatrix} \mathbf{u}_t^Z \\ \mathbf{u}_t^F \end{pmatrix},$$

where $\mathbf{Y}_t$ are the observable variables included in the DSGE model and $\mathbf{F}_t$ is a small vector of unobserved factors extracted from a large data-set of macroeconomic time series, that capture additional economic information relevant to model the dynamics of $\mathbf{Y}_t$. The system reduces to the standard VAR used to evaluate DSGE models if $\mathbf{\Phi}_{12}(L) = 0$.

Importantly, and differently from Boivin and Giannoni (2006), the FAVAR is not interpreted as the reduced form of a DSGE model at hand. In fact, in this case the restrictions implied by DSGE model on a general FAVAR are very difficult to trace and model evaluation becomes even more difficult to implement. A very tightly parameterized theory model can have a very highly parameterized reduced form if one is prepared to accept that the relevant theoretical concepts in the model are combination of many macroeconomic and financial variables.

The structural identification problem for DSGE has recently received some close attention (Canova and Sala (2006)). Moreover, in Consolo et al. (2007), they point up the importance of this concept in the comparison between the DSGE-VAR and DGSE-FAVAR by using diagnostic tests on residuals.

To the concept of mixture model can be applied the question concerning the statistical identification and there are two important contributions in creating a DSGE-FAVAR, compared to the DSGE-VAR approach.

In Consolo et al. (2007) and in Baurle (2008), they propose a method to combine the DSGE information with a FAVAR representation.

In Consolo et al. (2007), the VAR representation is augmented by using exogenous factors which have not an interpretation in the theoretical model. Instead, in Baurle (2008), the factors have an interpretation as variables from the DSGE model and they have a direct economic link (as in Boivin and Giannoni (2006)). In both papers, they find that the best forecasting performance is given by the combination between DSGE and FAVAR

In this sense, the Factor Augmented VAR is considered as a better alternative representation for the simple VAR. The use of FAVAR leads essentially to the best forecasting performance, since the FAVAR representation nests the VAR representation.

1.4 A Simple DSGE Model: the Approximation

One of the most important issue in the use of DSGE models in empirical macroeconomics is the statistical representation and the approximation of the theoretical model by using a statistical framework. In the macroeconomic literature of the last decade, DSGE models are used to explain the impact of a monetary policy shock on output and inflation, or the impact of a technology shock on labor hours. In the empirical evidence, the estimation are obtained by using Structural Vector Autoregressions (SVARs) in exercises of evaluation of DSGE models. In these applications, the structural parameters of a DSGE model are estimated by minimizing the distance between the model's and the estimated VAR impulse responses functions.

However, there is some debate whether SVARs can in practice discriminate between alternative DSGE models and whether their sampling properties are good enough to justify the popularity raised in the applied macroeconomics. In this criticism, Chari, Kehoe and McGrattan (2005) and Christiano, Eichenbaum and Vigfusson (2006) investigate the properties of the estimators based on SVAR by simulating an artificial data generating process (DGP) derived from a prototype real business cycle (RBC) model and by comparing true with estimated impulse responses. In both papers, they find that the estimates obtained via a long-run identification scheme are very imprecise and the failure of the use of finite-order SVAR is sometimes attributed to the fact that they are only approximations to VARMA / infinite-order VAR and it could happen that a finite VAR representation does not exist at all. In Cooley and Dwyer (1998), it is possible to find an example of an economic model that implies a vector autoregressive moving-average (VARMA) representation of the data series and state: "*While VARMA models involve additional estimation and identification issues, these complications do not justify systematically ignoring these moving average components, as in the SVAR approach*". It is obvious that a DSGE model implies restrictions in the mapping between economic shocks and observable variables. In linear models, these restrictions are summarized by Vector Moving Average (VMA). If the VMA representation is not invertible a DSGE model does not admit a VAR representation mapping economic shocks to a vector of observable variables and its lags. The problem of the approximation of the statistical representation of a DSGE model is given by the fact that the not all variables used in the model are observed. In this sense, the representation could not be a VAR described as in the previous section, but it is necessary to use a more complex representation, a VARMA format. In Fernandez-Villaverde et al. (2007), DSGE models generally imply a state space system that has a direct VARMA and eventually an infinite VAR representation. Moreover, in Fernandez-Villaverde et al. (2007) the big problem of the invertibility is discussed and some examples of well-specified DSGE models that lack a VAR representation are provided. Christiano et al. (2006) argue that "*The spec-*

ification error involved in using a finite-lag VAR is the reason that in some of our examples, the sum of VAR coefficients is difficult to estimate accurately". In Ravenna (2007), the main point in the representation issue is not only the existence of a VAR finite-order representation, but also the number of lags included in the VAR truncation. *Truncation bias* can affect the impulse response function through two separate channels: the VAR(p) erroneously constraints to zero some coefficients in the true VAR representation and the VAR(p) coefficients can lead to mistaken identification of the structural shocks, this effect is called *identification bias.*

Moreover, Chari et al. (2007) discuss that a VAR is not able to capture the underlying VARMA process by showing that the *truncation bias*, which can considered as the population bias resulting from applying a finite-order VAR, is the main source of the observed small sample bias in their simulation studies.

To show this problem of the approximation is possible to consider the previous model , but with not observable government and technology shock. There are as in the previous example 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:

$$\begin{aligned}
\widetilde{x_t} &= E_t[\widetilde{x_{t+1}}] - \frac{1}{\tau}(\widetilde{R_t} - E_t[\widetilde{\pi_{t+1}}]) + (1-\rho_G)\widetilde{g_t} + \rho_Z \frac{1}{\tau}\widetilde{z_t} && (1.12)\\
\widetilde{\pi_t} &= \beta E_t[\widetilde{\pi_{t+1}}] + \kappa[\widetilde{x_t} - \widetilde{g_t}] && (1.13)\\
\widetilde{R_t} &= \rho_R \widetilde{R_{t-1}} + (1-\rho_R)(\psi_1 \widetilde{\pi_t} + \psi_2 \widetilde{x_t}) + \epsilon_{R,t} && (1.14)\\
\widetilde{g_t} &= \rho_g \widetilde{g_{t-1}} + \epsilon_{g,t} && (1.15)\\
\widetilde{z_t} &= \rho_z \widetilde{z_{t-1}} + \epsilon_{z,t} && (1.16)
\end{aligned}$$

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.

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. 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 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 :

$$\Gamma_0 \widetilde{\mathbf{Z}}_t = \Gamma_1 \widetilde{\mathbf{Z}}_{t-1} + C + \Psi \epsilon_t + \Pi \eta_t \tag{1.17}$$

$t = 1, ..., T$ 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:

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

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

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

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

$$\widetilde{\mathbf{Z}}_t = \mathbf{T}(\theta)\,\widetilde{\mathbf{Z}}_{t-1} + \mathbf{R}(\theta)\,\varepsilon_t \tag{1.18}$$

$$\theta = [\kappa, \tau, \psi_1, \psi_2, \rho_R, \rho_g, \rho_Z, \sigma_R, \sigma_g, \sigma_Z]' \tag{1.19}$$

This representation is in a VARMA format and it can be truncated by using a finite-order VAR.

It is possible to evidence this problem of the approximation by using a general example. A linearized DSGE model can be written as a system of stochastic difference equations. The solution to the system is the recursive equilibrium law of motion (as in Ravenna, 2007):

$$y_t = Px_{t-1} + Qz_t \tag{1.20}$$

$$x_t = Rx_{t-1} + Sz_t \tag{1.21}$$

$$Z(L)z_t = \varepsilon_t \tag{1.22}$$

where x_t is an $n \times 1$ vector of endogenous state variables, z_t is an $m \times 1$ vector of exogenous state variables, y_t is an $r \times 1$ vector of endogenous variables, ε_t is a vector stochastic process of dimension $m \times 1$ such that $E(\varepsilon_t) = 0$, $E(\varepsilon_t \varepsilon_t') = \Sigma$, $E(\varepsilon_t \varepsilon_s') = 0$ for $t \neq s$ and Σ is a diagonal matrix. $Z(L)$ is the matrix polynomial $[I - Z_1 L - ... - Z_p L^p]$ in the lag operator L defining a stationary vector AR(p) stochastic process.

Ravenna (2007) stresses the possibility to obtain the results concerning the existence of a finite-order VAR representation by using the specification in equations (1.20) to (1.22). This specification of the equilibrium has two advantages. The first advantage regards the role of the variables in the finite-order VAR representation. It is possible that the endogenous and exogenous state vectors play a different role in the representation of the system by using a finite-order VAR, since these two kinds of variables have a different economic interpretation. Actually, an economic model is built to explain the dynamics of the endogenous components (y_t and x_t) and which are the observable variables; instead, the dynamics of the exogenous variables (z_t) is left unexplained by the model. The second advantage regards the importance of the matrix $Z(L)$ in subsequent results.

The polynomial $Z(L)$ is typically assumed to be of the first order. For example, in case of $Z(L) = [I - Z_1 L]$, it is possible to write the previous system of equations, defining the vector $\widetilde{x}_t = [x_{t-1}\ z_{t-1}]'$, as follows:

$$\begin{aligned}
y_t &= C\widetilde{x}_t + D\varepsilon_t && (1.23)\\
\widetilde{x}_{t+1} &= A\widetilde{x}_t + B\varepsilon_t && (1.24)\\
C &= [P\ QZ_1]; D = Q; A = \begin{bmatrix} R & SZ_1 \\ 0 & Z_1 \end{bmatrix}; B = \begin{bmatrix} S \\ I \end{bmatrix} && (1.25)
\end{aligned}$$

The matrices A, B, C and D depends on the parameter set of the model, θ, as in the specification proposed by Fernandez-Villaverde et al. (2007).

However, the conditions of the existence of a finite-order VAR representation can be explained by using a simple example, as discussed in Ravenna (2007).

It is possible to write a DSGE model solution as follows:

$$\begin{aligned}
Y_t &= AY_{t-1} + Bz_t && (1.26)\\
z_t &= Z_1 z_{t-1} + \varepsilon_t\\
Y_t &= \begin{bmatrix} x_t \\ y_t \end{bmatrix}; A = \begin{bmatrix} R & 0 \\ P & 0 \end{bmatrix}; B = \begin{bmatrix} S \\ Q \end{bmatrix}
\end{aligned}$$

The vector Y_t has dimension $1 \times n + r$ and all its components are observable; instead, the vector of unobservable variables, z_t has dimension $m = n + r$. Since the number $n + r$ of observable variables is equal to the number of shocks, if B^{-1} exists, $z_t = Z_1[B^{-1}Y_{t-1} - B^{-1}AY_{t-2}] + \varepsilon_t$.

Ravenna (2007) proposes a restricted VAR(2) representation for the previous system as:

$$\begin{aligned}
Y_t &= (A + BZ_1B^{-1})Y_{t-1} - (BZ_1B^{-1}A)Y_{t-2} + B\varepsilon_t\\
&= \Gamma_1 Y_{t-1} + \Gamma_2 Y_{t-2} + \eta_t
\end{aligned}$$

where the VAR innovations $\eta_t = B\varepsilon_t$ are a rotation of the structural shocks vector ε_t.

The condition of the existence of a VAR representation for this DSGE model is that $m = n + r$.

When $m > n + r$ a VAR representation of the DSGE model may exist, but it will not be possible to map η_t into a higher-dimension vector of orthogonal shocks ε_t.

When $m < n + r$, the system is singular, preventing likelihood estimation of the VAR.

It is possible to obtain a non-singular VAR representation of the model, dropping some of the observable variables from the system so as to satisfy the requirement $m = n+r$. Omitting $r - r_1$ rows of the y_t vector does not affect the VAR(2) representation of any other observable variable. Ravenna (2007) proposes a subset $\widehat{Y}_t$ of the vector Y_t, facing with the following conditions:

1. the vector x_t belongs to the set of observable variables included in the data sample;

2. the determinant of $\left[|G(L) + PD_G(L)SQ^{-1}L|\right]$ [2] is of degree zero in L.

If the first sufficient condition is not met, hence $\widehat{Y}_t$ has a finite order VARMA representation. Actually, as explained in Fernandez-Villaverde et al. (2007), the MA component is invertible and a VAR representation exists.

The second condition is necessary and sufficient to obtain a finite-order VAR representation.

The second crucial point in the VAR representation is the approximation in the number of lags. When conditions 1 or 2 are not met, a finite order VAR may still be a very good approximation to the true data generating process if the VAR matrix coefficients for longer lags of y_t are close to zero.

Estimating a DSGE model by using a VAR, it is common to include few lags in order to provide a reasonable approximation to the true VAR. This assumption can be misleading, since the truncation can affect the approximation of the VAR through two channels.

First, the truncated VAR coefficients are biased (*truncation bias)*: a VAR(p) does not describe the true dynamics of the DSGE model, since all coefficients for lags larger than p are restricted to be equal to zero.

[2] For this point, Ravenna (2007) uses the notation presented in eqs (1) to (3) and $G(L) = [I - RL]$ and $D_G(L)$ is the adjoint matrix of $G(L)$ (for more details, see Ravenna (2007)).

Second, if the VAR coefficients enter in the computation of the matrix identifying structural shocks from reduced form innovations, truncation results in an identification error (*identification bias*).

According to this part of the literature, it seems necessary to approximate carefully the VARMA representation of a theoretical model and in this sense it could be possible to use the standard and classical information criteria on the real data (such as Akaike, Schwartz, Likelihood Ratio, Final Prediction Error and Hannah-Quinn). O

It is obvious that in the model validation in the DSGE approach, using the mixture model the DSGE-VAR, the VARMA representation and truncation of the theoretical model is crucial. Essentially, the dummy observations depend on the solution of the theoretical model with restrictions and the representation of the model should be accurate as possible. In both VAR representations, for the actual data and the artificial data, the lag-length must be the same. Hence, it is not important only the existence of a VAR finite-order representation, but also it is important to recognize the number of lags. In Paccagnini (2009), it is shown how increasing the number of lags in VAR representation could lead a misspecified results. It seems that mispecifying the number of lags, the weight of the economic model increases, even if the model used for generating the dummy observation priors is not the data generating process, otherwise it is possible to check the lag length by using the classical information criteria on the artificial data generated by the theoretical model.

1.5 What is next?

The recent research in DSGE model evaluation points out the importance of the use of the hybrid model. There are several applications of DSGE-VAR to compare different DSGE models for closed or open economy. For example, in Liu et al. (2008), DSGE-VAR is used to forecast South African Economy, in Adjemian et al. (2008), DSGE-VAR is used in order to compare different optimal monetary policy and in Adolfson et al. (2008) and Lees et al. (2007), the hybrid model, DSGE-VAR, is used to evaluate open economy models. The main points are given by the representation of the economic model, the approximation and the statistical identification. In Kascha and Mertens (2007) and in McGrattan (2006), they explore the possible advantages of the structural VARMA and state space models that capture the full structure of the time series representation implied by DSGE models, while imposing minimal theoretical assumptions.

In Kascha and Mertens (2009), they simulate DSGE models and fit different reduced form models to recover the structural shocks using the same long-run identification strategy by using different estimations algorithms. Among the interesting results, they show that SVARs do not perform poorly in these simulation exercises since they are only finite-order approximations, but the "bad" performance of SVAR is most likely due to the fact the long-run identification approach is inappropriate with small samples. Kascha and Martens (2009) argue "*We point out some properties of the simulated DGPs that make it hard to identify structural shocks for any method. The main problem with long-run restrictions is therefore not one of working with a specific model class*".

However, Kascha and Martens (2009) is one contribution of the importance of the structural aspect in the analysis of the DSGE models.

Even if the use of VAR on the actual data combined with the BVAR implied by the restrictions from the theoretical model has obtained a great success as an alternative approach to SVARs, the economist are always interesting in the concept of the structural for the policy analysis.

However, the DSGE-VAR is also considered as an example of a structural approach (Theodoridis, 2007) since the BVAR on the artificial data is implied by the restrictions imposed considering the theoretical model. In this sense, in Sims (2008) there is an interesting comment on the concept of "structural". According to Sims, it is possible to approximate a DSGE model as a structural VARMA. A DSGE model is linearized and solved to eliminate expectational terms, for example as the following format:

$$y_t = A^* (\theta) + H^* (\theta) \varepsilon_t$$

in the solution the matrices ($A^* (\theta)$ and $H^* (\theta)$) obtained by the algorithm are in function of the model's behavioral parameters. The "structural" interpretation is explained by the fact that observed data are a function of the disturbances ε_t, whose interpretations are carried over from the DSGE. Considering the idea behind the hybrid models, the DSGE model together with the prior distributions for the parameter of the models (expressed by using vector θ) are used in order to generate a prior distribution for the matrices A and H, in the representation in the real world.

These matrices, A and H, has a distribution centered on ($A^{*}(\theta)$ and $H^{*}(\theta)$), the linearized coefficients of the solution of the model.

Sims proposes the use of dummy observation priors to recover the oscillation frequencies in the business cycle range that could be more heavily weighted. In this sense, the main point is the possibility to allow to weight the dummy observations more heavily at frequencies where the econometrician can give more credence to the DSGE model. The suggestion takes into consideration the eigenvalues and eigenvectors of the cross-moments from the theoretical model, getting rid of the eigenvalues smaller than a certain threshold, imposed arbitrarily, in order to consider only the long-run component of the restrictions.

Another excellent contribution in this way, but at this moment it is still in preliminary stage, is given by Del Negro, Diebold and Schorfheide (2008). They evidence the problem concerning the frequency by referring to the spectral analysis literature.

In this paper, they evidence that the misspecification of DSGE models is more prevalent at some frequencies than at others and they extend the DSGE-VAR framework by considering dummy observation priors from a DSGE model that have been transformed into the frequency domain. They explain that most of DSGE models are designed for business-cycle analysis and in general one often does not expect them capture high frequency or long-run movements in the data. They argument that the current generation of DSGE models is severely misspecified in terms of their low frequency implications. For example, a model as one proposed by Smets and Wouters (2007), a medium scale model applied for monetary policy analysis is not able to generate the persistence in the great ratios, in particular the consumption-output ratio or labor share, observed in the quarterly U.S. data. Moreover, as discussed in Whelen (2000) and Edge, Kiley and Laforte (2005), many of the great ratios are not stationary as implied by the standard DSGE model. Models that impose invalid long-run restrictions on the data tend to be quickly rejected against specifications that allow for a more general trend structure, such as VARs. In this way, it is possible to apply to the concept of DGSE-VAR the literature of the low frequency variation from the data prior to model estimation and evaluation, as in Watson (1993) and Diebold, Ohanian and Berkowitz (1998).

It is obvious that using the spectral analysis is possible to extract more information on the frequency problem, instead the approach proposed by Sims (2008) is easier, under a computational point of view, but less clear. However, these two contributions are an important development of the use of the dummy observations in the macroeconomics. Actually, from the use of "Minnesota priors" to the cross-moments from the DSGE model, there has been a great evolution in the use of the prior dummies (for an excellent discussion on dummy observation priors, see Sims (2005)). In the use of DSGE-VAR, there are not only the dummy observations which play a key role, but it is possible to take into consideration the issue concerning the approximation of an economic model by using a finite VAR or finite FAVAR. This representation is given by the solution of the economic model and it could be possible to consider a solution which embodies the structural component of the theoretical model. In this sense, the representation could be affected by the presence of the structural shocks. In other words, it is possible, as suggested by Del Negro et al. (2008) and Sims (2008), to overcome the problem of the representation of the theoretical model as a SVAR, but starting directly from the results shown by Ravenna (2007) and Fernandez-Villaverde et al. (2007), using a reduced form the model and recovering the structural aspect from the dummy observations. This approach could be considered not so clear and direct, but in this sense it is possible to select the restrictions imposed by the model by using a frequency analysis.

1.6 Concluding Remarks

In the recent literature, DSGE models are becoming very popular but what is very important is to understand how it is possible to estimate. The growing econometric technique allows the economist to estimate these models by using VAR representation and Bayesian VAR approaches. One important application is the hybrid or mixture model, the so-called DSGE-VAR which combines the VAR representation for the real data and the BVAR representation for the artificial data which are generated by the restrictions which come from the economic model. In this sense it is possible to realize a model validation in the DSGE approach. As the use of the VARs has been criticized considering the problems of the existence of a finite-order VAR representation and of the proper statistical representation, the DSGE-VAR is subjected to these problems.

The use of the mixture model, which is an hybrid format embodies the "structural" concept of DSGE models, has overcome the use of a mere SVARs representations. Even if in the recent macroeconometric literature, there are several examples of new "structural" applications by considering the DSGE-VAR, taking the long-run restrictions. It is possible to evidence the importance of new comments in order to develop the use of the dummy observation priors in the macroeconomic field and to exploit in the best way the estimation and the validation of the DSGE models, combining the actual data and the theoretical restrictions.

Bibliography

[1] Amisano, Gianni and Carlo Giannini (1996): "*Topics in Structural VAR Econometrics*", Springer–Verlag.

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

[3] 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.

[4] An, Sungbae, and Frank Schorfheide (2007): "Bayesian Analysis of DSGE Models", *Econometric Reviews,* 26 (2-3): pp1-60.

[5] Anderson, T.W. and Herman Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations", *Annals of Mathematical Statistics* 20, 46—63.

[6] Basmann, R.L. (1960): "On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistic", *Journal of the American Statistical Association* 55. 650—659.

[7] Bernanke, Ben S., Jean Boivin and Piotr Eliasz (2005): " Measuring the Effects of Monetary Policy a Factor-Augmented Vector Autoregressive (FAVAR) Approach", *The Quarterly Journal of Economics*, MIT Press, vol. 120(1), pages 387-422, January.

[8] Boivin, Jean and Marc P. Giannoni (2006): "DSGE Models in a Data-Rich Environment", NBER Working Papers 12772.

[9] Baurle, Gregor (2008): "Priors from DSGE Models for Dynamic Factor Analysis", Universitat Bern Discussion Paper, 08-03.

[10] Canova, Fabio (2007): "Methods for Applied Macroeconomic Research", Princeton University Press.

[11] Canova, Fabio and Luca Sala (2006): "Back to Square one: identification issues in DSGE models" , IGIER Working Paper 303, Università Bocconi.

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

[13] 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*.

[14] Christiano, Lawrence J., Martin Eichenbaum and Charles Evans (1998): "Monetary Policy Shocks: What Have We Learned and to What End?", NBER Working Paper No. 6400

[15] Christiano, Lawrence J., Martin Eichenbaum and Charles Evans (2005): "Nominal Rigidities and the Dynamic effects of a Shock to Monetary Policy", Journal of Political Economy 113, 1-45.

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

[17] Chudik, Alexander and M. Hashem Pesaran (2007): "Infinite Dimensional VARs and Factor Models", CESifo Working Paper 2176.

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

[19] Cogley, Timothy and Thomas J. Sargent (2005): "Drifts and Volatilities: Monetary Policies and Outcomes in post WWII U.S.", *Review of Economic Dynamics*, 8 (2), 262-302.

[20] Consolo, Agostino, Carlo A. Favero and Alessia Paccagnini (2007): "On the Statistical Identification of DSGE Models", IGIER Working Paper, 324, Bocconi University.

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

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

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

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

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

[26] Del Negro, Marco and Frank Schorfheide (2006): "How Good is what You've Got? DSGE-VAR as a Toolkit for evaluating DSGE Models", *Federal Reserve Bank of Atlanta Economic Review.*

[27] Del Negro, Marco, Frank Schorfheide, Frank Smets and Raf Wouters (2007): "On the Fit of New-Keynesisan Models", *Journal of Business, Economics and Statistics, 25,2, 124-162.*

[28] Del Negro, Marco, Francis X. Diebold and Frank Schorfheide (2008): "Priors from Frequency-Domain Dummy Observations", mimeo.

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

[30] Edge, Rochelle, Michael Kiley, and Jean-Philippe Laforte (2005): "An Estimated DSGE Model of the US Economy", Manuscript, Board of Governors.

[31] Fernandez-Villaverde Jesus, Juan Rubio-Ramirez, Thomas J. Sargent and Mark W. Watson (2007): "ABCs (and Ds) of understanding VARs", *the American Economic Review*, 97, 3,

[32] Fernandez-Villaverde Jesus (2009): "The Econometrics of DSGE Models", NBER Working Paper 14677.

[33] Giannoni, Domenico, Lucrezia Reichlin and Luca Sala (2005): "Monetary Policy in Real Time", CEPR Discussion Papers 4981, C.E.P.R. Discussion Papers.

[34] Geweke, John (1977): "The Dynamic Factor Analysis of Economic Time Series", in: D.J. Aigner and A.S. Goldberg (eds.), *Latent Variables in Socio-Economic Models*, North-Holland, Amsterdam.

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

[36] Hendry, David F. (1995): "*Dynamic Econometrics*", Oxford: Oxford University Press.

[37] 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.

[38] Kadiyala, K. Rao and Sune Karlsson (1997): "Numerical Methods for Estimation and Inference in Bayesian VAR-Models", *Journal of Applied Econometrics*, 12(2), 99-132.

[39] Kascha, Christian and Karel Mertens (2009): "Business Cycle Analysis and VARMA Models", *Journal of Economic Dynamics and Control,* Vol. 33 (2), pp. 267-282.

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

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

[42] Landon-Lane, John S. (1998): "Bayesian Comparison of Dynamic Macroeconomic Models", Ph.D. Dissertation, University of Minnesota.

[43] 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.

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

[45] Litterman, Robert B. (1986): "A Statistical Approach to Economic Forecasting", *Journal of Business and Statistics*, 4(1):1-4".

[46] 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.

[47] Lucas, Robert E. Jr. (1972): "Expectations and the Neutrality of Money", Journal of Economic Theory 4(2), 103-124.

[48] 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.

[49] Lucas, Robert E. Jr and Thomas J. Sargent (1979): "After Keynesian Macroeconomics", Federal Reserve Bank of Minneapolis, Quarterly Review, 3: 1-16.

[50] Maddala, G.S. (1988): "*Introduction to Econometrics*", Mac Millan, New York, NY.

[51] McGrattan, Ellen (2006): "Measurement with Minimal Theory", Federal Reserve Bank of Minneapolis Working Paper 643.

[52] Otrok, Christopher (2001): "On Measuring the Welfare Cost of Business Cycles", *Journal of Monetary Economics* 47, 61-92.

[53] Paccagnini, Alessia (2009): "Assessing the Potential of DSGE Model Evaluation in a Bayesian Framework", Manuscript, Università Bocconi.

[54] Ravenna Federico (2007): "Vector Autoregressions and Reduced Form Representations of DSGE models", *Journal of Monetary Economics*, 54, 7, 2048-2064.

[55] Ruge-Murcia, Francisco J. (2007): "Methods to Estimate Dynamic Stochastic General Equilibrium Models", *Journal of Economic Dynamics and Control* 31(8), 2599-2636.

[56] Sargent, Thomas J. and Christopher A. Sims (1977): "Business Cycle Modeling without Pretending to Have Too Much A-Priori Economic Theory", in C. Sims et al. (eds), *New Method in Business Cycle Research*, Federal Reserve Bank of Minneapolis, Minneapolis.

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

[58] Schorfheide, Frank (2008): "Bayesian Methods in Macroeconometrics", *The New Palgrave Dictionary of Economics*. Second Edition. Eds. Steven N. Durlauf and Lawrence E. Blume. Palgrave Macmillan.

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

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

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

[62] Sims, Christopher A. (2008): "Making Macro Models Behave Reasonably", manuscript, Princeton University.

[63] Sims, Christopher A. and Tao Zha (1998): "Bayesian Methods for Dynamic Multivariate Models", *International Economic Review*, 39, 949-968.

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

[65] Smets, Frank and Raf Wouters (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach", *American Economic Review*, American Economic Association, vol. 97(3), pages 586-606, June.

[66] Spanos, Aris (1990): "The Simultaneous-Equations Model Revisited: Statistical Adequacy and Identification", *Journal of Econometrics, 44, 87-105.*

[67] Stock, James and Mark Watson (1999) "Forecasting Inflation", *Journal of Monetary Economics,* Vol. 44, no. 2

[68] Stock, James and Mark Watson (2002): "Macroeconomic Forecasting Using Diffusion Indexes", *Journal of Business Economics and Statistics*, XX:II, 147-162.

[69] Stock, James and Mark Watson (2005): "Implications of Dynamic Factor Models for VAR Analysis", NBER Working Paper No. 11467.

[70] Theodoridis, Konstantinos (2007): "Dynamic Stochastic General equilibrium (DSGE) Priors for Bayesian Vector Autoregressive (BVAR) Models: DSGE Model Comparison", Cardiff Business School Working Paper Series, E2007/15.

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

[72] Whelan, Karl (2000): "Balanced Growth Revisited: A Two-Sector Model of Economic Growth", Manuscript, Board of Governors.

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

Chapter 2

On the Statistical Identification of DSGE Models

Abstract[1]

Dynamic Stochastic General Equilibrium (DSGE) models are now considered attractive by the profession not only from the theoretical perspective but also from an empirical standpoint. As a consequence of this development, methods for diagnosing the fit of these models are being proposed and implemented. In this article we illustrate how the concept of statistical identification, that was introduced and used by Spanos (1990) to criticize traditional evaluation methods of Cowles Commission models, could be relevant for DSGE models. We conclude that the recently proposed model evaluation method, based on the $DSGE - VAR(\lambda)$, might not satisfy the condition for statistical identification. However, our application also shows that the adoption of a FAVAR as a statistically identified benchmark leaves unaltered the support of the data for the DSGE model and that a DSGE-FAVAR can be an optimal forecasting model.

Keywords: Bayesian analysis; Dynamic stochastic general equilibrium model; Model evaluation,Statistical Identification,Vector autoregression, Factor-Augmented Vector Autoregression.

JEL Classification: C11, C52

[1]This chapter is a joint paper with Agostino Consolo and Carlo A. Favero.

2.1 Introduction

Dynamic Stochastic General Equilibrium (DSGE) models are now considered attractive by the profession not only from the theoretical perspective but also for empirical analysis and for econometric policy simulation.[2] Model evaluation is an issue of crucial importance before policy simulation. Therefore, methods for diagnosing the fit of these models are being proposed and implemented. This article illustrates how the concept of statistical identification, originally introduced to criticize traditional evaluation methods of Cowles Commission models, could also be applied to the diagnostic tools recently proposed for DSGE models.

The concept of statistical identification has been introduced by Spanos (1990). Structural models can be viewed statistically as a reparameterization, possibly (in case of over-identified models) with restrictions, of the reduced form. Spanos distinguishes between structural identification and statistical identification. Structural identification refers to the uniqueness of the structural parameters, as defined by the reparameterization and restriction mapping from the statistical parameters in the reduced form, while statistical identification refers to the choice of a well-defined statistical model as reduced form. Diagnostics for model evaluation are constructed in Cowles commission tradition in a way that is closely related to the solution of the identification problem. In fact, in the (very common) case of over-identified models, a test of the validity of the over-identifying restrictions can be constructed by comparing the restricted reduced form implied by the structural model with the reduced form implied by the just-identified model in which each endogenous variables depend on all exogenous variables with unrestricted coefficients. The statistics are derived in Anderson and Rubin (1949) and Basman (1960). The logic of the test attributes a central role to the structural model. The statistical model of reference for the evaluation of the structural model is derived by the structural model itself. Spanos (1990) points out that the root of the failure of the Cowles Commission approach lies in the little attention paid to the statistical model implicit in the estimated structure. Any identified structure that is estimated without checking that the implied statistical model is an accurate description of the data is bound to fail if the statistical model is not valid. The Spanos

[2] See An and Schorfheide(2006) and the JBES Invited address presented at the Joint Statistical Meeting 2006 "On the Fit of New Keynesian Models" by Del Negro, Schorfheide, Smets and Wouters, published on the April 2007 issue of the JBES with comments by L.Christiano, R.Gallant, C.Sims, J.Faust, and L.Kilian.

critique of the Cowles commission approach lies naturally within the LSE approach to econometric modelling. Such approach reverses the prominence of the structural model with respect to the reduced form representation. The LSE approach starts its specification and identification procedure with a general dynamic reduced form model. The congruency of such a model cannot be directly assessed against the true DGP, which is unobservable. However, model evaluation is made possible by applying the general principle that congruent models should feature true random residuals; hence, any departure of the vector of residuals from a random normal multivariate distribution should signal a mis-specification. A structural model can be identified and estimated only after a validation procedure based on a battery of tests on the reduced form residuals has been satisfactorily implemented. A just-identified specification does not require any further testing, as its implied reduced form does not impose any further restrictions on the baseline statistical model. The validity of over-identified specification is instead tested by evaluating the validity of the restrictions implicitly imposed on the general reduced form. Interestingly, the lack of statistical identification offers an explanation for the failure of the Cowles Commission models very different from the "great critiques" by Lucas (1976) and Sims (1980), that concentrate on model failure related to structural identification problems.

The structural identification problem for DSGE has recently received some close attention (Canova and Sala (2006)). This paper concentrates on the statistical identification model of DSGE models. We illustrate how the logic of some recently proposed model evaluation tools for DSGE models, based on the comparative evaluation of a DSGE-VAR model with an unrestricted VAR model, resembles closely the logic applied within the Cowles Commission approach in testing for the validity of over-identifying restrictions in structural models. We then show that statistical identification can be achieved by using a Factor Augmented VAR (FAVAR), and we compare the properties of DSGE-VAR and DSGE-FAVAR. We provide an empirical illustration by considering the case of a very simple three-equations DSGE model (Del Negro and Schorfheide (2004)).

2.2 Statistical Identification: the original concept

Spanos (1990) considers the case of a simple demand and supply model to show how the reduced form is ignored in the traditional approach. The example is based on the market for commercial loans discussed in Maddala (1988). Most of the widely used estimators allow the derivation of numerical values for the structural parameters without even seeing the statistical models represented by the reduced form. Following this tradition, the estimated (by 2SLS) structural model is a static model that relates the demand for loans to the average prime rate, to the Aaa corporate bond rate and to the industrial production index, while the supply of loans depends on the average prime rate, the three-month bill rate and total bank deposits. The quantity of commercial loans and the average prime rate are considered as endogenous while all other variables are taken as, at least, weakly exogenous variables in the sense of Engle et al. (1983) and no equation for them is explicitly estimated. Given that there are two omitted instruments in each equation, one over-identifying restriction is imposed in both the demand and supply equations. The validity of the restrictions is tested via the Anderson-Rubin (1949) tests, and leads to the rejection of the restrictions at the 5 per cent level in both equations, although in the second equation the restrictions cannot be rejected at the 1 per cent level. This mild evidence against the adopted structural model ignores the fact that estimation of the statistical model, i.e. the reduced form implied by the adopted identifying restrictions, yields a specification for which the underlying statistical assumptions of linearity, homoscedasticity, absence of autocorrelation and normality of residuals are all strongly rejected. On the basis of this evidence the adopted statistical model is not considered as appropriate. An alternative model allowing for a richer dynamic structure (two lags) in the reduced form is then considered. Such dynamic specification is shown to provide a much better statistical model for the data than the static reduced from. Of course, the adopted structural model implies many more over-identifying restrictions than the initial more parsimonious specification. When tested, the validity of these restrictions is overwhelmingly rejected for both the demand and the supply equations. Such evidence leads to the conclusion that the lack of statistical identification of the original model might lead to failure of rejecting the structural model of interest when it is false.

In practice, Cowles Commission models have been abandoned because of their empirical failure and because of the great critiques related to their lack of structural identification, much less emphasis has been posed by the mainstream literature on the problem of statistical identification, with the notable exception of the LSE approach to econometric dynamics (see, Hendry (1995)). Cowles Commission models for policy evaluation have been replaced by Dynamic Stochastic General Equilibrium (DSGE) models.

2.3 The Statistical Identification of VAR and DSGE models

The general linear (or linearized around equilibrium) DSGE model takes the following form(see Sims (2002)):

$$\mathbf{\Gamma}_0\mathbf{Z}_t = \mathbf{\Gamma}_1\mathbf{Z}_{t-1} + C + \Psi\epsilon_t \ + \Pi\eta_t \tag{2.1}$$

Where C is a vector of constants, ϵ_t is an exogenously evolving random disturbance, η_t is a vector of expectations errors, $\left(E_t\left(\eta_{t+1}\right) = \mathbf{0}\right)$, not given exogenously but to be treated as part of the model solution. The forcing processes here are the elements of the vector ϵ_t, this typically contains processes like Total Factor Productivity or policy variables that are not determined by an optimization process. Policy variables set by optimization, typically included $\mathbf{Z}_t$, are naturally endogenous as optimal policy requires some response to current and expected developments of the economy. Expectations at time t for some of the variables of the systems at time t+1 are also included in the vector $\mathbf{Z}_t$,whenever the model is forward looking. Model like (2.1)can be solved using standard numerical techniques (see, for example, Sims, 2002), and the solution can be expressed as:

$$\mathbf{Z}_t = \mathbf{A}_0 + \mathbf{A}_1\mathbf{Z}_{\mathbf{t-1}} + \mathbf{R}\epsilon_t \tag{2.2}$$

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

$$\mathbf{Z}_t = \mathbf{A}_0 + \mathbf{A}_1\mathbf{Z}_{\mathbf{t-1}} + \mathbf{u}_t \tag{2.3}$$

Within this framework a new role for empirical analysis based on reduced form models emerges, that is to provide evidence on the stylized facts to be matched by the theoretical model adopted for policy analysis and to decide between competing DSGE models. Given the estimation of a VAR the selection of a particular DSGE model among different alternatives can be based on the following three steps (see Christiano, Eichenbaum and Evans (1998)):

1. policy shocks of interests are identified in actual economies, i.e. in a VAR without theoretical restrictions;
2. the response of relevant economic variables to these shocks is then described;
3. finally, the same experiment is performed in the model economies (DSGE) to compare actual and model-based responses as an evaluation tool and a selection criterion for theoretical models.

The identification of the shocks of interest is the structural identification problem in VAR-based model evaluation. VAR modelling recognizes that identification and estimation of structural parameters is impossible without explicitly modelling expectations, therefore a structure like (2.3) can only be used to run special experiments that do not involve simulating different scenarios for the parameters of interests. A natural way to achieve these results is to experiment with the shocks ϵ_t. *Facts* are then provided by looking at impulse response analysis, variance decompositions and historical decompositions. All these experiments are run by keeping estimated parameters unaltered. Importantly, running these experiments is easier if shocks to the different variables included in the VAR are orthogonal to each other, otherwise it would not be possible to simulate a policy shock by maintaining all the other shocks at zero. As a consequence, VAR models need a structure because orthogonal shocks are normally not a feature of the statistical model. This fact generates the structural identification problem. The relation between (2.3) and (2.2)implies that:

$$\mathbf{u}_t = \mathbf{R}\epsilon_t,$$

from which we can derive the relation between the variance-covariance matrices of $\mathbf{u}_t$ (observed) and $\boldsymbol{\nu}_t$ (unobserved) as follows:

$$E\left(\mathbf{u}_t\mathbf{u}_t'\right) = \mathbf{R}E\left(\epsilon_t{\epsilon_t}'\right)\mathbf{R}'. \tag{2.4}$$

Substituting population moments with sample moments the relevant structural shocks can be identified and a number of parameters in the $\mathbf{R}$ matrix equal up to the number of different elements in the variance covariance matrix of the VAR innovations ($m(m+1)/2$, where m is the number of variables included in the VAR) can be estimated. As usual, for such a condition also to be sufficient for identification, no equation in (2.4) should be a linear combination of the other equations in the system (see Amisano and Giannini (1996), Hamilton (1994)). As for traditional models, we have the three possible cases of under-identification, just-identification and over-identification. The validity of over-identifying restrictions can be tested via a statistic distributed as a χ^2 with the number of degrees of freedom equal to the number of over-identifying restrictions. But again structural identification of a VAR is a totally different from statistical identification. In fact, misspecification of a VAR generated by omitted variables does not prevent structural identification but it leads to lack of statistical identification.[3]

Recent Model Evaluation of DSGE models exploits the fact that a solved RBC model is a statistical model. In fact, solved DSGE model often generates a restricted MA representation for the vector of observable variables of interest, that can be approximated by a VAR of finite order (see Fernandez-Villaverde et al.(2007) and Ravenna, (2007)). Interestingly, this recent approach to model evaluation does not require identification of structural shocks but it is still potentially affected by lack of statistical identification.

To see this point consider the general case of system (2.2) in which only a subset n of the m variables included in $\mathbf{Z}_t$ is observable and define such subset $\mathbf{Y}_t$. $\mathbf{Y}_t$ has a VAR(∞) representation. This is usually approximated by a finite VAR representation at the cost of a truncation that can be relevant for purposes such as the identification of structural shocks (see Ravenna (2007)). Note that if the RBC model features a number of shocks smaller than the number of variables included in the VAR, some of the VAR shocks are interpreted as measurement error.

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

[3]Think for example of a simple Data Generating Process made of a bivariate cointegrated VAR with two strucutral shocks: a temporary one and a permanent one. The estimation of a VAR in difference omitting the cointegrating relations would not prevent structural identification of the two shocks but it would certainly lead to lack of statistical identification.

$$\mathbf{Y}_t = \mathbf{\Phi}_0^*(\theta) + \mathbf{\Phi}_1^*(\theta)\mathbf{Y}_{t-1} + ... + \mathbf{\Phi}_p^*(\theta)\mathbf{Y}_{t-p} + \mathbf{u}_t^* \tag{2.5}$$

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

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

$$\underset{Txn}{\mathbf{u}^*} = \begin{bmatrix} \underset{Tx1}{\mathbf{u}_1^*} & ... & \underset{Tx1}{\mathbf{u}_n^*} \end{bmatrix} \tag{2.6}$$

$$\underset{Txn}{\mathbf{Y}} = \begin{bmatrix} \underset{Tx1}{\mathbf{Y}_1} & ... & \underset{Tx1}{\mathbf{Y}_n} \end{bmatrix} \tag{2.7}$$

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

$$\underset{1x(np+1)}{\mathbf{X}_t'} = \begin{bmatrix} 1, \underset{1xn}{\mathbf{Y}_{t-1}'} ... \underset{1xn}{\mathbf{Y}_{t-p}'} \end{bmatrix} \tag{2.8}$$

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

where all coefficients are convolutions of the structural parameters in the model included in the vector θ. Of course the theoretical model imposes some restrictions on the VAR, that can be tested by evaluating them against the unrestricted VAR. Note that the relevant statistical model is constructed exactly as in the Cowles Commission approach: the specification of the statistical model is totally driven by that of the structural model. In fact, the statistical model is obtained by solving the structural model and then by relaxing some restrictions. As a matter of fact when this procedure is followed variables omitted from the structural model are never included in the statistical model and statistical identification becomes a potentially relevant issue. In a series of papers Del Negro and Schorfheide (2004 and 2006) and Del Negro, Schorfheide, Smets and Wouters (2004) adopt this line of research to propose a Bayesian framework for model evaluation. This method tilts coefficient estimates of an unrestricted VAR toward the restriction implied by a DSGE model. The weight placed on the DSGE model is controlled by an hyperparameter called λ. This parameter takes values ranging from 0 (no-weight on the DSGE model) to ∞ (no weight on the unrestricted VAR). Therefore, the posterior distribution of λ provides an overall assessment of the validity of the DSGE model restrictions.

The chosen benchmark to evaluate this model is the unrestricted VAR derived from the solved DSGE model

$$\begin{aligned}
\mathbf{Y}_t &= \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{Y}_{t-1} + ... + \mathbf{\Phi}_p \mathbf{Y}_{t-p} + \mathbf{u}_t && (2.9) \\
\mathbf{u}_t &\sim N\left(\mathbf{0}, \Sigma_u\right) \\
\mathbf{Y} &= \mathbf{X\Phi} + \mathbf{u} \\
\underset{(np+1)xn}{\mathbf{\Phi}} &= \left[\underset{nx1}{\mathbf{\Phi}_0}, \underset{nxn}{\mathbf{\Phi}_1}, ..., \underset{nxn}{\mathbf{\Phi}_p}\right]', && (2.10)
\end{aligned}$$

where:

$$\begin{aligned}
\mathbf{\Phi} &= \mathbf{\Phi}^*\left(\theta\right) + \mathbf{\Phi}^{\Delta} \\
\Sigma_u &= \Sigma_u^*\left(\theta\right) + \Sigma_u^{\Delta}
\end{aligned}$$

the DSGE restrictions are imposed on the VAR by defining:

$$\begin{aligned}
\Gamma_{XX}\left(\theta\right) &= E_\theta^D\left[\mathbf{X}_t \mathbf{X}_t'\right] \\
\Gamma_{XZ}\left(\theta\right) &= E_\theta^D\left[\mathbf{X}_t \mathbf{Y}_t'\right]
\end{aligned}$$

where E_θ^D defines the expectation with respect to the distribution generated by the DSGE model, that of course have to be well defined. We then have:

$$\mathbf{\Phi}^*\left(\theta\right) = \Gamma_{XX}\left(\theta\right)^{-1}\Gamma_{XY}\left(\theta\right)$$

Beliefs about the DSGE model parameters θ and model misspecification matrices $\mathbf{\Phi}^{\Delta}$ and Σ_u^{Δ} are summarized in prior distributions, that, as shown in Del Negro and Schorfheide (2004) can be transformed into prior for the VAR parameters $\mathbf{\Phi}$ and Σ_u.In particular we have:

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

where the parameter λ controls the degree of model misspecification with respect to the VAR: for small values of λ the discrepancy between the VAR and the DSGE-VAR is large and a sizeable distance is generated between unrestricted VAR and DSGE estimators, large values of λ correspond to small model misspecification and for $\lambda = \infty$ beliefs about DSGE mis-specification degenerate to a point mass at zero. Bayesian estimation could be interpreted as estimation based a sample in which data are augmented by an hypothetical sample in which observations are generated by the DSGE model, within this framework λ determines the length of the hypothetical sample.

Given the prior distribution, posterior are derived by the Bayes theorem:

$$
\begin{aligned}
\Sigma_u \,|\theta, Y &\sim IW\left((\lambda+1) T \hat{\Sigma}_{u,b}(\theta), (\lambda+1) T - k, n\right) \\
\mathbf{\Phi}\,|\Sigma_u, \theta, Y &\sim N\left(\hat{\mathbf{\Phi}}_b(\theta), \Sigma_u \otimes \left[\lambda T \Gamma_{XX}(\theta) + \mathbf{X}'\mathbf{X}\right]^{-1}\right) \\
\hat{\mathbf{\Phi}}_b(\theta) &= \left(\lambda T \Gamma_{XX}(\theta) + \mathbf{X}'\mathbf{X}\right)^{-1}\left(\lambda T \Gamma_{XY}(\theta) + \mathbf{X}'\mathbf{Y}\right) \\
\hat{\Sigma}_{u,b}(\theta) &= \frac{1}{(\lambda+1) T}\left[\left(\lambda T \Gamma_{YY}(\theta) + \mathbf{Y}'\mathbf{Y}\right) - \left(\lambda T \Gamma_{XY}(\theta) + \mathbf{X}'\mathbf{Y}\right)\hat{\mathbf{\Phi}}_b(\theta)\right]
\end{aligned}
$$

which shows that the smaller λ, the closer the estimates are to the OLS estimates of an unrestricted VAR, the higher λ the closer the estimates are to the values implied by the DSGE model parameters θ.

In practice, a grid search is conducted on a range of values for λ to choose that value that maximize the marginal data density. The typical results obtained when using DSGE-VAR(λ) to evaluate models with frictions is that " ... the degree of misspecification in large-scale DSGE models is no longer so large as to prevent their use in day-to-day policy analysis, yet is not small enough that it cannot be ignored...".

DSGE-VAR model evaluation ignores the issue of specification of the statistical model. Although the models are different, the evaluation strategy in the DSGE-VAR approach is very similar to the approach of evaluating models by testing over-identifying restrictions without assessing the statistical model implemented in Cowles foundation models. In the Cowles Commission approach the statistical model was derived by taking the reduced form obtained by removing all the over-identifying restrictions implied by the theoretical model. As a consequence, as in the case of the market for loans analyzed by Spanos (1991), a static theoretical model would always be evaluated against a static statistical model, which would certainly lack statistical identification whenever the true Data Generating Process featured any form of dynamics. In the DSGE-VAR approach the fact that the a DSGE model is a restricted VAR is used to derive the statistical model by removing the DSGE restrictions from the VAR. However, the question of the validity of this unrestricted VAR to represent the data is not addressed. In fact, the DSGE-VAR approach is looser than the Cowles foundation approach: model based restrictions are not imposed and tested but are made fuzzy by imposing a distribution on them and then the relevant question becomes what is the amount of uncertainty that we have to add to model based restrictions in order to make them compatible with a model-derived unrestricted VAR representation of the data. In fact such representation might not represent the data. The natural question here is how well does this procedure do in rejecting false models? Spanos (1991) has shown clearly that generalizing the statistical model by adding features that are not included in the theoretical model (the omitted dynamics in the case of the static demand and supply for loans model considered in the original illustrative example) could lead to dramatic changes in the outcome of tests for over-identifying restrictions. Is the Spanos criticisms of Cowles Commission model evaluation applicable to DSGE-VAR model evaluation?

2.4 The statistical identification of a DSGE-FAVAR

A DSGE model is a restricted VAR, removing the restrictions from the VAR and using an unrestricted VAR as a statistical model would imply that the Spanos criticism of the Cowles Commission model evaluation approach is applicable to DSGE-VAR model evaluation if the unrestricted VAR cannot be considered statistically identified. If the unrestricted VAR is not statistically identified we might not be able to conduct a proper statistical evaluation of the theoretical model by using it as a benchmark. The important point made by Spanos in his contribution is that reduced form of theoretical must be evaluated against specifications that captures the relevant information in the data potentially omitted from the theoretical model. Obviously the omission of dynamics used by Spanos to criticize the statistical identification of static Cowles Commission models is not applicable to VAR specifications. However, there are a number of potential sources of mis-specification for the model derived VAR. Think of all those variables that are omitted from the theoretical model because of its specific nature, say fiscal policy in a model designed to analyze the effect of monetary policy or foreign variables and the exchange rate in a closed economy model, but also of all variables that are not theory related but are relevant to determine the actual behavior of policy makers. A good example is the commodity price index and the problem of modelling of the behavior of monetary policy authority. Early VARs for the analysis of monetary policy that did not include in the information set a commodity price index tended to deliver a "price puzzle", i.e. a positive response of prices to an unexpected monetary tightening. Such anomaly has been attributed to the existence of a leading indicator for inflation to which the Fed reacts and which is omitted from the VAR. The omission form the information set of a variable positively correlated with inflation and interest rates makes the VAR mis-specified and explains the positive relation between prices and interest rates observed in the impulse response functions. It has been observed (see Christiano, Eichenbaum and Evans (1998)) that the inclusion of a Commodity Price Index in the VAR solves the 'price puzzle'. DSGE model do not typically include the commodity price index in their specification as a consequence the VAR derived by relaxing the theoretical restrictions in a DSGE model is misspecified. So the evaluation of the effects of conducting model misspecification with a "wrong" benchmark is a practically relevant one.

As a matter of fact DSGE model tend to produce a high number of very persistent shocks (see Smets and Wouters (2003)), such evidence would have been certainly taken as a signal of model mis-specification by an LSE type methodology. Still the model do not do too badly when judged in the metric of the λ test.

Another dimension potentially relevant for evaluating the statistical model underlying DSGE-VAR is structural stability of the VAR parameters. If the DSGE restrictions are valid, then parameters in the VAR are convolutions of structural parameters that, by their nature, should be constant over time. Cogley and Sargent (2005), Primiceri (2005) and Sims and Zha (2006) point out to instability of reduced form VARs. Interestingly, the evidence is mixed on the source of instability: Cogley and Sargent (2005) and Primiceri (2005) point toward time variation in parameters and in disturbance variance, while, according to Sims and Zha (2006) the best fit allows time variation in disturbance variances only. Justiniano and Primiceri (2007) and Fernandez-Villaverde and Rubio Ramirez (2007a, 2007b) point out to parameter instability of DSGE models, and the evidence on the sources of instability is similar to that available for reduced form VAR, with Justiniano and Primiceri (2007) highlighting the importance of stochastic volatilities while Fernandez and Rubio Ramirez (2007b) questioning the stability of parameters deemed to be structural.

In the light of this evidence it would be important to consider as a benchmark for model evaluation a more general specification than the one obtained by releasing some coefficient restrictions on a VAR involving only the variables included in a DSGE model. Ideally, we would like to consider as a benchmark a model that parsimoniously includes all the information excluded from the theoretical DSGE. A recent strand of the econometric literature[4] has shown that very large macroeconomic datasets can be properly modelled using dynamic factor models, where the factors can be considered as an exhaustive summary of the information in the data. The rationale underlying dynamic factor models is that the behavior of several variables is driven by few common forces, the factors, plus idiosyncratic shocks. Hence, the factors can provide an exhaustive summary of the information in large datasets, and in this sense they are precious to alleviate omitted variable problems in empirical analysis using traditional small-scale models, see Bernanke and Boivin (2003). In fact, Bernanke and Boivin (2003), Bernanke, Boivin and Eliasz (2005) proposed to exploit factors in the estimation of VAR to generate a more general specification. Chudik and Pesaran (2007) illustrate how a VAR augmented by factor could help in keeping the number of parameters to be estimated under control without loosing relevant information. Boivin and Giannoni (2006) propose a DSGE-FAVAR as a way of removing the assumption that economic variables included in a DSGE are properly measured by a single indicator. The theoretical concepts of the model are treated as partially observed to use the information set in factors to map them. We shall use a factor-augmented VAR (FAVAR)

[4] Stock and Watson (2002), Forni and Reichlin (1996, 1998) and Forni et al. (1999, 2000)

as the relevant statistical model to conduct model evaluation. A FAVAR benchmark for the evaluation of a DSGE model will take the following specification:

$$\begin{pmatrix} \mathbf{Y}_t \\ \mathbf{F}_t \end{pmatrix} = \begin{bmatrix} \mathbf{\Phi}_{11}(L) & \mathbf{\Phi}_{12}(L) \\ \mathbf{\Phi}_{21}(L) & \mathbf{\Phi}_{22}(L) \end{bmatrix} \begin{pmatrix} \mathbf{Y}_{t-1} \\ \mathbf{F}_{t-1} \end{pmatrix} + \begin{pmatrix} \mathbf{u}_t^Z \\ \mathbf{u}_t^F \end{pmatrix},$$

where $\mathbf{Y}_t$ are the observable variables included in the DSGE model and $\mathbf{F}_t$ is a small vector of unobserved factors extracted from a large data-set of macroeconomic time series, that capture additional economic information relevant to model the dynamics of $\mathbf{Y}$

$_t$. The system reduces to the standard VAR used to evaluate DSGE models if $\mathbf{\Phi}_{12}(L) = 0$, therefore, within this context, the relevant λ test would add to the usual DSGE model-related restrictions on $\mathbf{\Phi}_{11}(L)$ the restrictions $\mathbf{\Phi}_{12}(L) = 0$.To our knowledge, FAVAR have not been so far used to evaluate DSGE, and this is what we shall do in this paper using dynamic factors as the analogue of a richer dynamics for the evaluation of Cowles commission models proposed by Spanos[5].

Importantly, and differently from Boivin and Giannoni (2006), we do not interpret the FAVAR as the reduced form of a DSGE model at hand. In fact, in this case the restrictions implied by DSGE model on a general FAVAR are very difficult to trace and model evaluation becomes even more difficult to implement. A very tightly parameterized theory model can have a very highly parameterized reduced form if one is prepared to accept that the relevant theoretical concepts in the model are combination of many macroeconomic and financial variables. Identification of the relevant structural parameters, that is very hard also in DSGE model with observed variables (see Canova and Sala (2006)), becomes even harder. Natural advantages of this approach are increased efficiency in the estimation of the model and improved forecasting performance. However, model evaluation becomes almost impossible to pursue and a theoretical model can only by rejected by another theoretical model, while the implied statistical model is made so general that virtually no room is left to the data to reject a DSGE model.

[5] In our application we consider a special case of the FAVAR in which $\Phi_{21}(L) = 0$

2.5 Model Evaluation of a Simple DSGE Model

We consider a small New Keynesian DSGE model of the economy which features a representative household optimizing over consumption, real money holdings and leisure, a continuum of monopolistically competitive firms with price adjustment costs and a monetary policy authority which sets the interest rate. Furthermore, the model is driven by three exogenous processes which determine government spending, g_t, the stationary component of technology, z_t, and the policy shock, $\epsilon_{R,t}$.

A full description of the model can be found in Woodford (2003). Here, we mainly focus on its log-linear representation which takes each variable as deviations from its trend. The model has a deterministic steady state with respect to the de-trended variables: the common component is generated by a stochastic trend in the exogenous process for technology. The model follows Del Negro and Schorfheide (2004) (henceforth, DS) and it reads

$$\tilde{x}_t = E_t\tilde{x}_{t+1} - \frac{1}{\tau}(\tilde{R}_t - E_t\tilde{\pi}_{t+1}) + (1-\rho_G)\tilde{g}_t + \rho_z\frac{1}{\tau}\tilde{z}_t \tag{2.11}$$

$$\tilde{\pi}_t = \beta E_t\tilde{\pi}_{t+1} + \kappa\,(\tilde{x}_t - \tilde{g}_t) \tag{2.12}$$

$$\tilde{R}_t = \rho_R\tilde{R}_{t-1} + (1-\rho_R)(\psi_1\tilde{\pi}_t + \psi_2\tilde{x}_t) + \epsilon_{R,t} \tag{2.13}$$

$$\tilde{g}_t = \rho_g\tilde{g}_{t-1} + \epsilon_{g,t} \tag{2.14}$$

$$\tilde{z}_t = \rho_z\tilde{z}_{t-1} + \epsilon_{z,t} \tag{2.15}$$

where $\tilde{x}_t$ is the output gap, $\tilde{\pi}_t$ is the inflation rate, $\tilde{R}_t$ is the short-term interest rate and $\tilde{g}_t$ and $\tilde{z}_t$ are two AR(1) stationary processes for government and technology, respectively.

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 up to 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 $\tilde{g}_t$. The parameter $0 < \beta < 1$ is the households' discount factor and $\tau > 0$ is the inverse of the elasticity of intertemporal substitution. The second equation is the forward-looking Phillips curve which describes the dynamics of inflation and κ determines the degree of the short-run trade-off between output and inflation.

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 shock $\epsilon_{R,t}$ can be interpreted as unanticipated deviation from the policy rule or as policy implementation error. The set of structural shocks is thus $\epsilon_t = (\epsilon_{R,t}, \epsilon_{g,t}, \epsilon_{z,t})'$ which collects technology, government and monetary shocks.

The model needs to be solved and this can be done by applying the algorithm proposed by Sims (2002). Define the vector of variables a $\tilde{Z}_t = \left(\tilde{x}_t, \tilde{\pi}_t, \tilde{R}_t, \tilde{R}_t^*, \tilde{g}_t, \tilde{z}_t, E_t\tilde{x}_{t+1}, E_t\tilde{\pi}_{t+1}\right)$ and the vector of shocks as $\epsilon_t = (\epsilon_{R,t}, \epsilon_{g,t}, \epsilon_{z,t})$. Therefore the previous set of equations, (2.11) - (2.15), can be recasted into a set of matrices $(\Gamma_0, \Gamma_1, C, \Psi, \Pi)$ accordingly to the definition of the vectors $\tilde{Z}_t$ and ϵ_t

$$\Gamma_0 \tilde{Z}_t = C + \Gamma_1 \tilde{Z}_{t-1} + \Psi\epsilon_t + \Pi\eta_t \tag{2.16}$$

where η_{t+1}, such that $E_t\eta_{t+1} \equiv E_t\left(y_{t+1} - E_t y_{t+1}\right) = 0$, is the expectations error[6].

As a solution to (2.16), we obtain the following policy function

$$\tilde{Z}_t = T\left(\theta\right)\tilde{Z}_{t-1} + R\left(\theta\right)\epsilon_t \tag{2.17}$$

and in order to provide the mapping between the observable data and those computed as deviations from the steady state of the model we set the following measurement equations as in DS

$$\Delta \ln x_t = \ln\gamma + \Delta\tilde{x}_t + \tilde{z}_t \tag{2.18}$$

$$\Delta \ln P_t = \ln\pi^* + \tilde{\pi}_t \tag{2.19}$$

$$\ln R_t = 4[(\ln R^* + \ln\pi^*) + \tilde{R}_t] \tag{2.20}$$

which can be also cast into matrices as

$$Y_t = \Lambda_0\left(\theta\right) + \Lambda_1\left(\theta\right)\tilde{Z}_t + v_t \tag{2.21}$$

where $Y_t = (\Delta\ln x_t, \Delta\ln P_t, \ln R_t)'$, $v_t = 0$ and Λ_0 and Λ_1 are defined accordingly. For com-

[6] See Appendix A for a detailed derivation.

pleteness, we write the matrices T, R, Λ_0 and Λ_1 as a function of the structural parameters in the model, $\theta = \left(\ln \gamma, \ln \pi^*, \ln r^*, \kappa, \tau, \psi_1, \psi_2, \rho_R, \rho_g, \rho_Z, \sigma_R, \sigma_g, \sigma_Z\right)'$: such a formulation derives from the rational expectations solution.

The evolution of the variables of interest, Y_t, is therefore determined by (2.17) and (2.21) which impose a set of restrictions across the parameters on the moving average (MA) representation. Given that the MA representation can be very closely approximated by a finite order VAR representation, DS propose to evaluate the DSGE model by assessing the validity of the restrictions imposed by such a model with respect to an unrestricted VAR representation. The choice of the variables to be included in the VAR is however completely driven by those entering in the DSGE model regardless of the statistical goodness of the unrestricted VAR.

The model evaluation method proposed by DS is based on a mixed estimation which combines the data information with the prior information deriving from the DSGE model. The source for the data information is the unrestricted VAR process for all variables included in the DSGE model. The measurement, (2.21), and the transition, (2.17), equation can be used to derive a sample of artificial data which are theory driven. This is effectively a set of dummy observations which can be added to the observable data as in Sims and Zha (1998)[7] to derive a prior distribution for the VAR coefficients. Furthermore, such a prior would be conjugate and that is relevant to keep tractability of the posterior analysis.

A further step would be to compute the posterior distribution for (Φ, Σ_e, θ). Such a posterior can be written as

$$P\left(\Phi, \Sigma_e, \theta \mid Y\right) = P_\Phi\left(\Phi, \Sigma_e \mid \theta, Y\right) \times P_\theta\left(\theta \mid Y\right), \tag{2.22}$$

where the first component can be easily calculated by using the conjugacy property of the DSGE-based prior while the second one, $P\left(\theta \mid Y\right)$, will be derived by recalling MCMC methods. In particular, following DS, the Metropolis-Hastings will be employed to approximate the posterior. The posterior mean of the estimated coefficients is governed by the parameter λ, a tightness parameters which determines the weight of the DSGE model in the posterior estimates. For instance, as $\lambda \to 0$, which means no artificial data from the DSGE, the posterior estimates will be equal to the maximum likelihood estimates of the unrestricted VAR since the prior would be

[7] We follow DS and work with population moments instead of artificial data generated from the restricted VAR(1) to avoid stochastic variation.

flat.[8] Alternatively, as $\lambda \to \infty$ we have the posterior driven by the DSGE model only. Therefore the λ that maximizes the data density is a natural criterion to assess the DSGE model against the VAR benchmark.[9] To sum up model evaluation requires to simulate from the posterior distribution of θ and this results is obtained by going through a number of steps[10]:

1. Set a value of λ or assume a discrete grid over which to run the computation;
2. Solve the DSGE model and get the population moments used in the prior;
3. Given λ, find the posterior moments for (Φ, Σ_e) and the marginal data density $P(Y \mid \theta)$;
4. Construct the kernel of the posterior for θ, $P_Y(Y \mid \theta) \times P(\theta)$;
5. Apply the MH acceptance method in order to generate a Markov chain from the posterior distribution of θ;
6. By applying the Gelfand-Dey (1994) method, with the correction proposed by Geweke (1999), compute the marginal data density of the model for each λ;
7. Compare such marginal densities over the discrete grid of λ. Model validation requires λ that maximizes the data density.

2.6 DSGE Model Evaluation and Statistical Identification.

The evaluation of DSGE models based on the λ parameter is based on the choice of a VAR derived by relaxing the theoretical restrictions as a statistical benchmark. This choice closely resemble the approach taken by the Cowles Commission to evaluate structural econometric models: the chosen benchmark, being driven the specification of the structural model adopted, could very well lack of statistical identification.

To evaluate the potential relevance of this problem we propose to base the evaluation of the DSGE model on a model-independent benchmark, which is based on a larger information set than the VAR driven by the DSGE model specification.

[8] The Jeffrey's prior we used for the DSGE based prior.
[9] See Appendix C for a detailed description of how draws from the posterior distribution are generated.
[10] Details of the derivation of the relevant posterior are described in Appendix C.

We consider the case in which additional economic information, not fully captured by $\mathbf{Y}_t$, is relevant to modelling the dynamics of inflation output growth and the monetary policy rate. These additional information can be summarized in a (small) $(kx1)$ vector of unobserved factors $\mathbf{F}_t$.

We then adopt a Factor Augmented VAR as our benchmark model:

$$\begin{pmatrix} \mathbf{Y}_t \\ \mathbf{F}_t \end{pmatrix} = \begin{bmatrix} \mathbf{\Phi}_{11}(L) & \mathbf{\Phi}_{12}(L) \\ \mathbf{\Phi}_{21}(L) & \mathbf{\Phi}_{22}(L) \end{bmatrix} \begin{pmatrix} \mathbf{Y}_{t-1} \\ \mathbf{F}_{t-1} \end{pmatrix} + \begin{pmatrix} \mathbf{u}_t^Z \\ \mathbf{u}_t^F \end{pmatrix},$$

The system reduces to the standard VAR used to evaluate DSGE models if $\mathbf{\Phi}_{12}(L) = 0$, therefore, within this context, the relevant λ test would add to the usual DSGE model-related restrictions on $\mathbf{\Phi}_{11}(L)$ the restrictions $\mathbf{\Phi}_{12}(L) = 0$.

The implementation of the Bayesian framework described for the evaluation of the DSGE model is altered only as far the likelihood function is concerned, where the more general FAVAR specification substitutes the VAR model (2.3) .

Factors can be constructed following a very recent strand of the econometric literature which has shown that very large macroeconomic datasets can be properly modelled using dynamic factor models, where the factors can be considered as an exhaustive summary of the information in the data.

We extract factors from "informational" time series included in $(Nx1)$ vector X_t,that consists of a balanced panel of 131 monthly macroeconomic time-series (updates of the series used in Stock and Watson (2002)). The number of informational time series N is large (larger than time period T) and must be greater than the number of factors and observed variables in the FAVAR system $(k + M \ll N)$.

We estimate our FAVAR by implementing a two-step estimation (Bernanke, Boivin and Eliasz (2005)).

We assume that the informational time series X_t are related to the unobservable factors F_t by the following observation equation:

$$X_t = \Lambda^f \mathbf{F}_t + e_t \tag{2.23}$$

where $\mathbf{F}_t$ is a $r \times 1$ vector of common factors, Λ^f is a (Nxk) matrix of factor loadings, Λ^y is

(NxM) and the $(Nx1)$ vector of error terms e_t are mean zero and are normal and uncorrelated or with a small cross-correlation, in fact, the estimator we employ allows for some cross-correlation in e_t that must vanish as N goes to infinity. Note that this representation nests also models where X_t depends on lagged values of the factors, see Stock-Watson(2002) for details.

In the first step factors are obtained from the observation equation by imposing the orthogonality restriction $F'F/T = I$.This implies that $\widehat{F} = \sqrt{T}\widehat{G}$, where the $\widehat{G}$ are the eigenvectors corresponding to the K largest eigenvalues of XX', sorted in descending order. Stock and Watson (2002) showed that the factors can be consistently estimated by the first r principal components of X, even in the presence of moderate changes in the loading matrix Λ. For this result to hold it is important that the estimated number of factors, k, is larger or equal than the true number, r.

In the second step, we estimate the FAVAR equation replacing $\mathbf{F}_t$ by $\widehat{\mathbf{F}_t}$.We shall then compare the VAR and the FAVAR and complete the analysis by considering a DSGE-VAR and a DSGE-FAVAR.

The standard VAR adopted as a benchmark to assess DSGE models is a nested model into FAVAR structure. The FAVAR structure is a richer specification than parsimoniously summarizes a much larger information set than that considered in the VAR.

We shall use the FAVAR for evaluating the statistical identification of the VAR by taking several steps.

First, we shall assess directly the significance of coefficient on factors and compare the goodness of fit of the FAVAR with respect to that of the VAR. We shall also evaluate how different is impulse response analysis based on the VAR and on the FAVAR to see how different is the description of the economy offered by the two alternative models.

Second, the two alternative models will be analyzed by assessing via appropriate tests, as suggested by Spanos (1990), the properties of homoskedasticity, serial correlation and normality of the residuals.

Third, the out-of-sample forecasting performance of the alternative models will be assessed by evaluating the RMSE of the FAVAR, the VAR, and the DSGE to assess the relevance of the information progressively discarded by the different models in forecasting the macroeconomic variables of interest.

Finally, the DSGE-FAVAR will be used as a benchmark for the implementation of the lambda test proposed by Del Negro-Schorfheide (2004) to assess how the optimal lambda is influenced by the choice of the FAVAR rather than the VAR as a statistical model to be combined with the DSGE.

2.7 Empirical Results

2.7.1 The Data

We analyse the DSGE-VAR model proposed by Del Negro and Schorfheide (2004) (DS). DS based their analysis on U.S. quarterly data from 1955:III to 2001:III, they also analyse separately the 1955-1979 period and the 1980-2001 period. We concentrate on the second subsample to keep comparability between our DSGE-FAVAR and the original DS DSGE-VAR whilst using a sample period which has experienced a more stable monetary and financial structure and a lesser volatility of macroeconomic variables period. Structural breaks in mean and volatility are found in the literature by comparing the pre-80 with the post-80 period, while the null hypothesis of parameters' stability cannot be rejected in the post 80 period (see Justiniano and Primiceri (2007)), moreover inflation, the monetary policy rate and annual real output growth, i.e. all variables included in the empirical specification, are clearly mean reverting in the post 1980 period. This evidence should reduce the concern of having a non-stationary VAR that omits potential long-run cointegrating relations among the variables of interest. 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 lag length in the VAR is four quarters. In order to construct the FAVAR we proceed to extract factors from a balanced panel of 131 monthly macroeconomic and financial time series (Stock and Watson (1999)) The dataset involves several measures of industrial production, interest rates,various price indices, employment as well as other important macroeconomic and also financial variables. This panel data is in monthly format, we transform it into a quarterly dataset using end-of-period observations. All series have been transformed to induce stationarity. The series are taken into level, logarithms, first or second difference (in level or logarithms) according to series characteristics (see the Appendix for a description of all series and details of the transformations). Following Bernanke, Boivin and Eliasz (2005) we partition the data in two categories of information variables: slow and fast. The partitioning is crucial to identify shocks necessary to construct impulse response functions in our FAVAR. Slow-moving variables (for example, wages or spending) do not respond contemporaneously to unanticipated changes in monetary policy; while fast-moving variables (for example, asset prices and interest rates) do respond contemporaneously to monetary shocks (see again the Appendix for further details).

We proceed to extract two factors from slow variables and one factor from fast variables and we call them respectively "slow factors" and "fast factor". [11] We extract factors by using principal components. Stock-Watson (1998) showed that the factors can be consistently estimated by the first r principal components of X, even in the presence of moderate changes in the loading matrix Λ. For this result to hold it is important that the estimated number of factors, k, is larger or equal than the true number, r. Bai and Ng (2000) proposed a set of selection criteria to choose k that are generalizations of the BIC and AIC criteria. As suggested by Bai and Ng (2000) we use information criteria to determine the number of factors but, as they are not so decisive, we limit the number of factors to three to strike a balance between the variance of the original series explained by the principal components and the difference in the parameterization of the VAR and the FAVAR. It is also worth noting that the factors are not uniquely identified, but this is not a problem in our context because we will not attempt a structural interpretation of the estimated factors. Finally, having determined the number of factors, we specify a Factor Augmented VAR by considering four-lags of the factors to keep the same lag-order chosen by DS for the VAR, we also consider a more parsimonious parameterization in which only one-lag of the factors is included.

[11] We extract factors by using principal components. As suggested by Bai and Ng(2000) we use information criteria to determine the number of factors but, as they are not so decisive, we limit the number of factors to three to strike a balance between the variance of the original series explained by the principal components and the difference in the parameterization of the VAR and the FAVAR.

2.7.2 The DSGE-VAR

We consider a benchmark DSGE-VAR model that replicates the results reported in Del Negro and Schorfheide (2004). As discussed in the previous section we report estimates for the DSGE-VAR over the sample 1981-2001, considering the DSGE model described in section 2 and a four-order VAR for the vector $Y_t = (\Delta \ln x_t, \Delta \ln P_t, \ln R_t)'$. We report in Table 1 reports Prior and Posterior for DSGE model parameters that are calibrated to generate posterior means and intervals as in Table 2 in Del Negro and Schorfheide (2004).

TABLE 1: Prior and Posterior of DSGE Model Parameters 1981-2001

	Prior		Posterior		Posterior		Posterior		Posterior	
			($\lambda = 0.2$)		($\lambda^* = 0.6$)		($\lambda = 1$)		($\lambda = 10$)	
	LOW	UPP	LOW	UPP	LOW	UPP	LOW	UPP	LOW	UPP
$\ln \gamma$	0.101	0.922	0.314	0.923	0.378	0.926	0.388	0.914	0.440	0.859
$\ln \pi^*$	0.219	1.863	0.511	1.112	0.503	1.080	0.474	1.087	0.288	1.548
$\ln r^*$	0.132	0.880	0.144	0.746	0.186	0.757	0.234	0.789	0.500	0.866
κ	0.063	0.513	0.144	0.701	0.198	0.804	0.236	0.820	0.062	0.405
τ	1.197	2.788	1.167	2.674	1.170	2.475	1.114	2.604	2.005	3.601
ψ_1	1.121	1.910	1.010	1.643	1.005	1.522	1.000	1.539	0.999	1.366
ψ_2	0.001	0.260	0.111	0.524	0.165	0.699	0.174	0.663	0.240	0.617
ρ_R	0.157	0.812	0.402	0.791	0.488	0.756	0.530	0.751	0.723	0.837

Notes: LOW and UPP are the lower and the upper bounds of the 90% confidence intervals based on the output of the Metropolis-Hastings Algorithm.

We then conduct DSGE model evaluation by determining $\widehat{\lambda}$ using the same grid presented in Del Negro and Schorfheide (2004), $\Lambda = \{0.20, 0.60, 1, 1.4, 1.8, 10, Inf\}$.

The minimum value of λ satisfying the lower bound restriction $\lambda \geq \frac{k+m}{T}$ with $k = 13$, $m = 3$ and $T = 80$ is $\lambda_{\min} = .20$. Figure 1 reports the results of the grid search that deliver 0.60 as the optimal λ in case we use Metropolis-Hastings Algorithm 100 000 replications[12].

Insert Figure 1 here

[12] Slightly different results are obtained when using 25000 replications, as the mapping between lambda and the marginal data density is not as smooth as with 100000 replications.

Note that the weight attached to the DSGE is $\frac{\lambda}{1+\lambda}$ so $\lambda^* = .60$ implies a weight of 0.375 on the DSGE model and therefore the size of the artificial sample generated by the DSGE should be of sixty per cent of the size of the sample of genuine observations. On the basis of very similar evidence Del Negro, Schorfheide, Smets and Wouters (2006) conclude that "...the degree of misspecification in DSGE models is no longer so large to prevent their use in day-to-day policy analysis, yet it is not so small that it cannot be ignored....".

2.7.3 The Statistical Identification of the DSGE-VAR

We begin our assessment of the statistical identification of the VAR used to construct the DSGE-VAR model by illustrating the statistical evidence on the augmentation of the VAR with factors.

In practice, we consider the extension of the baseline VAR model:

$$\begin{aligned}\mathbf{Y}_t &= \sum_{i=1}^{4} A_i Y_{t-i} + \mathbf{u}_t^Y \\ Y_t &= (\Delta \ln x_t, \Delta \ln P_t, \ln R_t)\end{aligned}$$

to the following FAVAR model

$$\begin{aligned}\begin{pmatrix} \mathbf{Y}_t \\ \mathbf{F}_t \end{pmatrix} &= \begin{bmatrix} \mathbf{\Phi}_{11}(L) & \mathbf{\Phi}_{12}(L) \\ \mathbf{\Phi}_{21}(L) & \mathbf{\Phi}_{22}(L) \end{bmatrix} \begin{pmatrix} \mathbf{Y}_{t-1} \\ \mathbf{F}_{t-1} \end{pmatrix} + \begin{pmatrix} \mathbf{u}_t^Y \\ \mathbf{u}_t^F \end{pmatrix} \\ \mathbf{Y}_t &= (\Delta \ln x_t, \Delta \ln P_t, \ln R_t) \\ \mathbf{F}_t &= \left(F_{1t}^s, F_{2t}^s, F_{3t}^f\right)\end{aligned}$$

where F_{1t}^s, F_{2t}^s are the two slow factors and F_{3t}^f is the fast factor. $\mathbf{\Phi}_{11}(L), \mathbf{\Phi}_{12}(L), \mathbf{\Phi}_{22}(L)$ are polynomial of order four in the lag factor for our benchmark parameterization. We experiment also with having $\mathbf{\Phi}_{12}(L), \mathbf{\Phi}_{22}(L)$ as polynomial of order one.

Table 2 compares the VAR and FAVAR specifications for the vector $\mathbf{Y}_t = (\Delta \ln x_t, \Delta \ln P_t, \ln R_t)$, considering two alternative FAVARs' including respectively one lag (FAVAR(1)) and four lags (FAVAR(4)) of the factors .

TABLE 2: VAR and FAVAR specifications: 1981-2001				
Equation		$\Delta \ln x_t$	$\Delta \ln P_t$	$\ln R_t$
	Adj R^2	0.39	0.30	0.93
VAR	S.E.	0.54	0.32	0.69
	Adj R^2	0.39	0.43	0.98
FAVAR(4)	S.E.	0.54	0.29	0.46
	$\chi^2\,(12)$	13.05 0.36	27.88 0.006	99.77 0.000
	Adj R^2	0.47	0.44	0.97
FAVAR(1)	S.E.	0.50	0.28	0.47
	$\chi^2\,(3)$	14.02 0.002	20.08 0.0002	83.94 0.000

The results reported in Table 2 clearly illustrate that factors are jointly significant in the specification for all three variables included in the baseline VAR, the only exception being the specification for the output growth equation when four lags of three factors are considered.

Table 3.1-3.3 report the evidence on the residual analysis from the VAR, the FAVAR(1) and the FAVAR(4). Table 3.1 reports the outcome of the Jarque-Bera (1980) tests of the null hypothesis of normality of residuals from each equation and for the joint three–equation model.

TABLE 3.1: Normality of Residuals					
Equation		$\Delta \ln x_t$	$\Delta \ln P_t$	$\ln R_t$	Joint
	Jarque-Bera	$\chi^2(2)$	$\chi^2(2)$	$\chi^2(2)$	$\chi^2(6)$
VAR		5.72 0.08	5.03 0.06	0.40 0.82	11.17 0.08
FAVAR(4)		10.51 0.005	2.52 0.28	8.76 0.01	21.79 0.001
FAVAR(1)		6.48 0.04	1.13 0.56	3.81 0.14	11.44 0.08

The null of normality is not rejected for the VAR and FAVAR(1) while it is rejected in the case of the FAVAR(4), the main cause of this rejection is the non-normality of residuals in the output growth equation. However, departure from the null hypothesis of normality of the size described by Table 3 has been shown to be very little relevant for the Bayesian analysis of the optimal λ,(see Christiano(2007)).

Table 3.2 reports the outcome of Breusch-Godfrey[13] Lagrange Multiplier test for autocor-

[13] See Godfrey(1988).

relation of residuals at all lags from one to four.

TABLE 3.2: Serial Correlation of Residuals				
LM $\chi^2(9)$	LAG 1	LAG 2	LAG 3	LAG 4
VAR	31.43 0.0002	29.37 0.0006	8.58 0.48	7.28 0.60
FAVAR(4)	11.97 0.21	8.44 0.49	13.06 0.16	12.77 0.17
FAVAR(1)	11.67 0.23	15.14 0.09	10.17 0.34	6.43 0.69

Here the results points toward strong evidence of residual autocorrelation in the VAR specification while the null hypothesis of absence of residual correlation at any lags cannot be rejected in the FAVAR(1) and the FAVAR(4) specifications.

TABLE 3.3: Homoscedasticity of Residuals			
White test	VAR	FAVAR(4)	FAVAR(1)
	$\chi^2(144)$	$\chi^2(288)$	$\chi^2(180)$
	172 0.05	290 0.44	208 0.08

Table 3.3 reports the outcome of the White (1980) heteroscedasticity tests on the residuals of the trivariate system. Once again while the null of homoscedasticity cannot be rejected in the FAVAR(4) and the FAVAR(1) specification, it is rejected at the five per cent level in the VAR specification. To further assess the stability of the variance-covariance matrix of residuals in our specifications we have estimated a diagonal BEKK (Engle and Kroner (1995)) allowing for a multivariate GARCH in the vector of residuals. Defining as Σ_t the (potentially) time-varying variance-covariace of residuals, BEKK is defined as:

$$\Sigma_t = \Omega\Omega' + A\mathbf{u}_{t-1}\mathbf{u}'_{t-1}A' + B\Sigma_{t-1}B$$

where $\mathbf{u}_{t-1}$ is the vector of residuals from the relevant model and the A, B are restricted to be diagonals. In the case of the FAVARs there are no significant coefficients in the A matrices, while these coefficients are significant in the VAR specification[14]. This evidence confirms the

[14] Results of the BEKK estimation are available upon request

findings in Giannone et al.(2008) that find that time-varying pattern in volatility of the shocks is a feature of small system that might disappear when the information set is enlarged.

We proceed to a further comparative analysis of the VAR and the FAVAR models by considering impulse response function to a monetary policy shock. Monetary policy shocks are identified in the VAR by assuming that the macroeconomic variables, inflation and output growth, take at least one period before responding to monetary policy while monetary policy is allowed to react simultaneously to macroeconomic variables. In the FAVAR identification is achieved by extending the VAR assumptions for macroeconomic variables and interest rates to slow factor and by assuming that the fast factor responds contemporaneously to all other variables in the system and that monetary policy does not contemporaneously react to the fast factor.

We plot in Figure 2 we plot responses for an horizon of 20 periods of quarterly inflation, quarterly output growth and the Federal Fund Rates to a monetary shock as derived in the VAR and in the FAVAR(4) estimated over the usual sample impulse in case of VAR and FAVAR for the usual sample 1981-2001. We also report one-standard deviation confidence intervals for the VAR estimation.

Insert Figure 2 here

The impulse responses show virtually no difference between the VAR and the FAVAR in the case of output growth, while in there are some differences in the case of inflation and the Federal Fund. In the case of inflation the FAVAR does not deliver the initial "price puzzle" that is observed with VAR based impulse responses and the negative dynamic response of inflation to a restrictive monetary policy at the one-year horizon is much more pronounced in the FAVAR case. In the case of the Federal Fund rate a much less persistent profile is observed in the FAVAR specification.

We complete our traditional evaluation of alternative models by considering the out-of-sample forecasting performance of the VAR, the FAVAR and the DSGE models. Given estimation of all models over the sample 1981:1-1997:4, we consider the out-of-sample performance for the period 1998:1-2001:4. In particular, we concentrate on the Root Mean Squared Error of the forecasting errors from the different model, computed as follows:

$$
\begin{aligned}
RMSE^{y} &= \sqrt{\frac{1}{16}\sum_{h=1}^{16}\left(y_{t+h}-\hat{y}_{t+h|t}\right)^{2}} \quad (2.24)\\
y &= \Delta\ln x_t, \Delta\ln P_t, \ln R_t,\\
t &= 1997:4
\end{aligned}
$$

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

We report the results of our analysis in Table 4.

TABLE 4: The Forecasting Performance of alternative models

MODEL		$\Delta\ln x_t$	$\Delta\ln P_t$	$\ln R_t$
		$RMSE$	$RMSE$	$RMSE$
VAR(4)		0.63	0.29	0.88
FAVAR(4,4)		0.56 (0.89)	0.24 (0.82)	0.92 (1.05)
FAVAR(4,1)		0.57 (0.92)	0.24 (0.82)	0.83 (0.94)
DSGE		0.63 (1.01)	0.24 (0.83)	0.80 (0.91)
DSGE-VAR($\lambda^* = 0.6$)		0.61 (0.97)	0.25 (0.86)	0.80 (0.91)

RMSE relative to the VAR(4) within brackets

FAVAR(4,i) includes i lags of the factors

Our results clearly favour the FAVAR against the VAR, moreover the improvements in the forecasting performance achieved by the DSGE and the DSGE-VAR ($\lambda^* = 0.6$) against the VAR are not obtained when the FAVARs are considered as benchmarks.

2.7.4 A FAVAR Analysis of the Simple DSGE Model

In the light of the evidence reported in the previous section it seems interesting to apply the mixed estimation technique to evaluate the properties of the DSGE-FAVAR instead of the DSGE-VAR. The FAVAR has the interesting properties of being an empirical model that is based on information independent from the theoretical model and it does then constitute a model whose statistical identification is independent of the validity of unrestricted VAR underlying the solution of the adopted theoretical model. In fact, we have shown for our particular application that a FAVAR which augments the VAR(p) specification for the variables in the theoretical model with a set of factors extracted from a large information set improves considerably on the VAR in terms of statistical adequacy.

In this case the benchmark specification for the unrestricted dynamics of the variables included in the theoretical model becomes the following:

$$Y_t = B_0 X_t + B_1 F_t + E_t \tag{2.25}$$

where $\mathbf{Y}_t = (\Delta \ln x_t, \Delta \ln P_t, \ln R_t)\,,\ X_t = [1, \mathbf{Y}_{t-1}, ...\mathbf{Y}_{t-p}]\,,\ F_t = \left[f_t', f_{t-1}', \ldots f_{t-q}'\right]'$ groups q lags of the three factors $f_t = [f_{1,t}, f_{2,t}, f_{3,t}]'$ extracted and interpreted as in Bernanke, Boivin and Eliasz (2005), E_t is the three-variate vector of innovations. System (2.59) can be re-written in a more compact form as follows:

$$Y_t = BW_t + E_t \tag{2.26}$$

where $B = [B_0, B_1]$ is of dimension $m \times (1 + mp + rq)$ and $W_t = [X_t', F_t']'$.

At this stage the derivation of the likelihood function resembles very closely the simpler case discussed in section 3. However, there are some differences in terms of the prior and the posterior distribution between the DSGE-VAR and the DSGE-FAVAR. We report in Appendix B the technical discussion of these derivation. Posterior calculations are similar to those discussed in the case of the DSGE-VAR, however in this case the parameter λ captures the relative weight of the information coming from the FAVAR and from the theoretical model.

The parameter λ is chosen from an interval which is unbounded from above. In our empirical exercise we will be using a discrete grid over which we will compute the marginal data density, $P(Y \mid \lambda)$. The minimum value, $\lambda_{\min} = \frac{m+k}{T}$, is model dependent and it is related to the existence of a well-defined Inverse-Wishart distribution. For completeness, it is worth to mention that $\lambda = 0$ refers to the FAVAR model with no prior and it is not possible to compute the marginal likelihood in this particular case. Therefore, we can show the marginal data density for any value of λ larger than $\lambda_{\min}$.Importantly $\lambda_{\min}$ depends on the degrees of freedom in the FAVAR and therefore, given estimation on the same number of available observation, $\lambda_{\min}$ for a DSGE-FAVAR will always be larger that $\lambda_{\min}$ for a DSGE-VAR.

Figure 3 shows the marginal likelihood for different λ, when a FAVAR(4,1) is considered as the baseline statistical model. The optimal value turn out to be $\lambda^* = 0.60$, as in the case of the DSGE-VAR. Of course, the distance between the optimal λ and $\lambda_{\min}$ is smaller in the DSGE-FAVAR than in the DSGE-VAR but still the lambda test indicates that the size of the artificial sample generated by the DSGE should be of sixty per cent of the size of the sample of genuine observations generated from the FAVAR model. In the case of a FAVAR(4,4), $\lambda^* = 1.4$ and the size of the artificial sample generated by the DSGE should now be greater than the size of the sample of genuine observations generated from the FAVAR model. Also in this case $\lambda_{\min}$ is higher than in our benchmark case as a consequence of the more generous parameterization of the DSGE-FAVAR(4,4).

To provide further evidence of the performance of the DSGE evaluated on the basis of the FAVAR, Table 5 considers the Forecasting performance of the VAR, the FAVAR and the optimal combination between DSGE and FAVAR.

TABLE 5: The Forecasting Performance of FAVAR and DSGE-FAVAR

MODEL		$\Delta \ln x_t$	$\Delta \ln P_t$	$\ln R_t$
		$RMSE$	$RMSE$	$RMSE$
VAR(4)		0.63	0.29	0.88
FAVAR(4,4)		0.56 (0.89)	0.24 (0.82)	0.92 (1.05)
FAVAR(4,1)		0.57 (0.92)	0.24 (0.82)	0.83 (0.94)
DSGE-FAVAR(4,4)($\lambda^* = 1.4$)		0.55 (0.88)	0.23 (0.79)	0.75 (0.85)
DSGE-FAVAR(4,1)($\lambda^* = 0.6$)		0.58 (0.93)	0.24 (0.80)	0.76 (0.87)

RMSE relative to the VAR(4) within brackets

The evidence reported shows that best forecasting performance is achieved by the optimal combination of the DSGE and the FAVAR. As final evaluation criterion, consistent with the Bayesian estimation procedure, we have considered the odds ratio between the DSGE-VAR and the DSGE FAVAR. The odds ratio compares marginal data density associated to alternative models, the traditional rule of thumb suggest to favour a model with respect to a competitor if its marginal data density is larger than three times that of its competitor. In our case the log of marginal likelihood of the DSGE-VAR in correspondence of the optimum λ is -215 while that of the DSGE-FAVAR(4,1) at the same λ is of -197.5. This implies that the odds ratio in log terms take the value of 17.5, clearly favouring the DSGE-FAVAR. Similar results are obtained by comparing the DSGE-FAVAR(4,4) with the DSGE-VAR. Importantly the comparison between Figure 1 and Figure 3 suggests that the dominance of the DSGE-FAVAR over the DSGE-VAR in terms of the posterior odds ratio is not limited to the optimal value of λ but it is independent from the choice of a particular value for λ. Overall, our results suggest that using a more general statistical model than that derived simply by relaxing restrictions from the solved theoretical model is important along two dimensions. First, it allows a further evaluation of the DSGE model against a larger information set. Second, in the case some support for the DSGE model is found in the data when evaluated against the larger information set (the optimal λ in the DSGE-FAVAR is different from zero), the optimal combination between the DSGE model and the statistical model based on a larger information set (the FAVAR) delivers a forecasting model (the DSGE-FAVAR) that dominates all alternatives.

2.8 Conclusions

In this paper we have analyzed the statistical identification of DSGE models by assessing if an unrestricted VAR constructed by relaxing cross-equation restrictions on the autoregressive approximation to the solution of a DSGE model is an appropriate statistical model. We have considered, as an alternative to the VAR, a FAVAR that uses a few factors to incorporate in the statistical model all the macroeconomic and financial information left out of the DSGE model.

Our application shows that, FAVAR models dominate VAR specification generated by adopting unrestricted version of the solution of DSGE models. Such dominance is clearly established by analysis of residuals and evaluation of forecasting performance. When we proceed to evaluate DSGE using FAVAR rather than VAR as statistical benchmark we find that some support for the DSGE model is still found in the data (the optimal λ in the DSGE-FAVAR is different from zero). Moreover, the optimal combination between the DSGE model and the statistical model based on a larger information set (the FAVAR) delivers a forecasting model (the DSGE-FAVAR) that dominates all alternatives.

The fact that the forecasting performance of the DSGE-FAVAR is the best among all alternatives, is somewhat reassuring against the worry that an artificially high value for the parameter λ might be chosen by maximizing the marginal likelihood. In fact, such criterion puts a considerable weight in favour of parsimony of specification, therefore more richly parameterized models might be unduly penalized by the lambda-test when they are evaluated against very parsimoniously parameterized theoretical models.

Our comparative analysis of the DSGE-VAR and the DSGE-FAVAR reiterates the point made by Christiano (2007) on the importance of complementing the value of the optimal λ with a cutoff function giving some weight to the difference between the number of free parameters in the unrestricted chosen statistical benchmark and in the DSGE model.

We conclude that the criticism of the Cowles Commission approach to model evaluation originally proposed by Spanos (1990) and centered on their lack of statistical identification might well apply to DSGE models and the recently proposed model evaluation method, based on the $DSGE-VAR(\lambda)$, is unlikely to detect the importance of such problem.

However, our application also shows that the adoption of a FAVAR as benchmark leaves unaltered the support of the data for the DSGE model and that a DSGE-FAVAR is the optimal forecasting model.

Bibliography

[1] Amisano, Gianni and Carlo Giannini (1996): "*Topics in Structural VAR Econometrics*", Springer–Verlag.

[2] An, Sungbae and Frank Schorfheide (2007): "Bayesian Analysis of DSGE Models", *Econometric Reviews,* 26 (2-3): pp1-60*.*

[3] Anderson, T.W. and Herman Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations", *Annals of Mathematical Statistics* 20, 46—63.

[4] Bai, Jushan and Serena Ng (2000): "Determining the Number of Factors in Approximate Factor Models", *Econometrica*, 70.

[5] Basmann, R.L. (1960): "On finite sample distributions of generalized classical linear identifiability test statistic", *Journal of the American Statistical Association* 55. 650—659.

[6] Bernanke, Ben S. and Jean Boivin (2003): "Monetary Policy in a Data-Rich Environment", *Journal of Monetary Economics*, L, 52-5464.

[7] Bernanke, Ben S., Jean Boivin and Piotr Eliasz (2005): " Measuring the Effects of Monetary Policy a Factor-Augmented Vector Autoregressive (FAVAR) Approach", *The Quarterly Journal of Economics*, MIT Press, vol. 120(1), pages 387-422, January.

[8] Boivin, Jean, and Marc P. Giannoni (2006):"DSGE Models in a Data-Rich Environment", NBER Working Papers 12772.

[9] Canova, Fabio and Luca Sala (2006):"Back to Square one: identification issues in DSGE models", IGIER Working Paper 303, Università Bocconi.

[10] 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.

[11] Christiano, Lawrence J., Martin Eichenbaum and Charles Evans (1998): "Monetary Policy Shocks: What Have We Learned and to What End?", NBER Working Paper No. 6400

[12] Christiano, Lawrence J., Martin Eichenbaum and Charles Evans (2005): "Nominal Rigidities and the Dynamic effects of a Shock to Monetary Policy", *Journal of Political Economy*, 113, 1-45.

[13] Chudik, Alexander and M. Hashem Pesaran (2007): "Infinite Dimensional VARs and Factor Models", CESifo Working Paper 2176.

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

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

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

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

[18] Del Negro, Marco, Frank Schorfheide, Frank Smets and Raf Wouters (2006): "On the Fit of New-Keynesisan Models", *Journal of Business, Economics and Statistics, 25,2, 124-162.*

[19] Engle Robert , David J. Hendry and J.F. Richard (1983): "Exogeneity", *Econometrica*, 51, 277-302.

[20] Engle Robert. and K.F. Kroner (1995): "Multivariate Simultaneous Generalized ARCH", *Econometric Theory*, 11, 122-150.

[21] Fernandez-Villaverde Jesus, Juan Rubio-Ramirez, Thomas J. Sargent and Mark W. Watson (2007): "ABCs (and Ds) of Understanding VARs", *the American Economic Review*, 97, 3.

[22] Fernandez-Villaverde J. and J.Rubio-Ramirez (2007a): "Estimating Macroeconomic Models: A Likelihood Approach", *Review of Economic Studies,* Volume 74, Number 4, October 2007 , pp. 1059-1087.

[23] Fernandez-Villaverde J. and J.Rubio-Ramirez (2007b): "How Structural are Structural Parameters?", NBER Macroeconomics Annual Working Paper 13166.

[24] Forni, Mario and Lucrezia Reichlin (1996): "Dynamic Common Factors in Large Cross-Sections", *Empirical Economics*, 21, 27-42.

[25] Forni, Mario and Lucrezia Reichlin (1998): "Let's Get Real: A Dynamic Factor Analytical Approach to Disaggregated Business Cycle", *Review of Economic Studies*, 65, 453-474.

[26] Forni, Mario, Marc Hallin, Marco Lippi and Lucrezia. Reichlin (1999): "Reference Cycles: The NBER Methodology Revisited", mimeo.

[27] Forni, Mario, Marc Hallin, Marco Lippi and Lucrezia. Reichlin (2000): "The Generalized Dynamic-Factor Model: Identification And Estimation,", *The Review of Economic and Statistics*, MIT Press, vol. 82(4), pages 540-554, November.

[28] Gelfand A.E. and D.K. Dey(1994) "Bayesian Model Choice: Asymptotics and Exact Calculations", *Journal of The Royal Statistical Society* Series B 56:501-514

[29] Geweke, John (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communications", *Econometrics Reviews*, 18, 1-127.

[30] Giannone Domenico, Michele Lenza and Lucrezia Reichlin (2008): "Explaining the Great Moderation. It is not the Shocks", ECB Working Paper Series, 865.

[31] Godfrey Leslie G. (1988): "*Misspecification Tests in Econometrics*", Cambridge University Press.

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

[33] Hamilton James D. (1994): "*Time Series Analysis*", Princeton University Press.

[34] Hendry, David F. (1995): "*Dynamic Econometrics*", Oxford: Oxford University Press.

[35] Jarque Carlos M. and Anil K. Bera (1980): "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals", *Economic Letters*, 6, 255-259.

[36] Justiniano, Alejandro and Giorgio Primiceri (2008): "The Time Varying Volatility of Macroeconomic Fluctuations", *The American Economic Review*, forthcoming

[37] 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.

[38] Maddala, G.S. (1988): "*Introduction to Econometrics*", Mac Millan, New York, NY.

[39] Primiceri, Giorgio (2005): "Time Varying Vector Autoregressions and Monetary Policy", *The Review of Economic Studies*, 72, 821-852

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

[41] Ravenna Federico (2007): "Vector Autoregressions and Reduced Form Representations of DSGE models", *Journal of Monetary Economics*, 54, 7, 2048-2064.

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

[43] Sims, Christopher A. (1992): "Interpreting the Macroeconomic Time Series facts: The Effects of Monetary Policy," *European Economic Review*, XXXVI, 975-1000.

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

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

[46] Sims, Christopher A. and Tao Zha (1998): "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, 39, 949-968.

[47] Sims, Christopher A. and Tao Zha (2006) "Were There Regime Switches in U.S. Monetary Policy?", *The American Economic Review*, 96(1), 54-81.

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

[49] Spanos, Aris (1990): "The Simultaneous-Equations Model Revisited: Statistical Adequacy and Identification", *Journal of Econometrics, 44, 87-105.*

[50] Stock, James and Mark Watson (1998): "Diffusion indexes", NBER WP 6702.

[51] Stock, James and Mark Watson (1999): "Forecasting Inflation", *Journal of Monetary Economics*, Vol. 44, no. 2

[52] Stock, James and Mark Watson (2002): "Macroeconomic Forecasting Using Diffusion Indexes", *Journal of Business Economics and Statistics*, XX:II, 147-162.

[53] White, Halbert (1980): "A Heteroscedastic-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity", *Econometrica,* 48, 817-838.

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

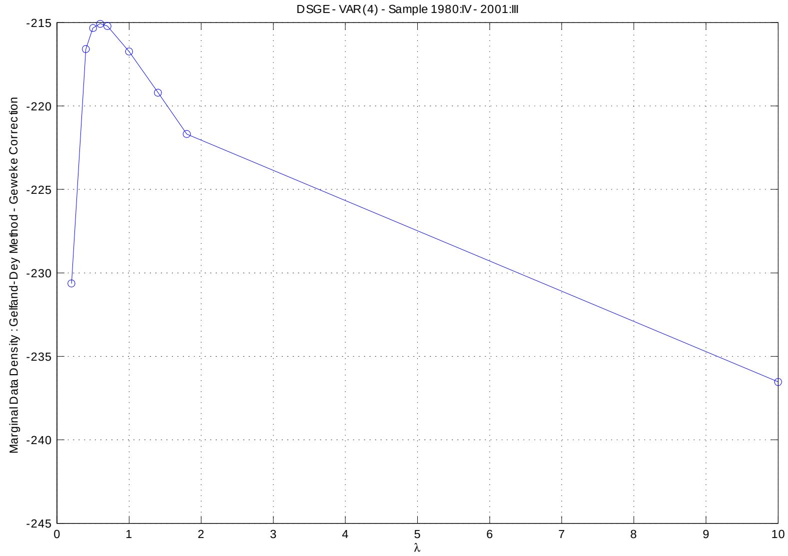

Figure 1: The optimal λ in the DSGE-VAR

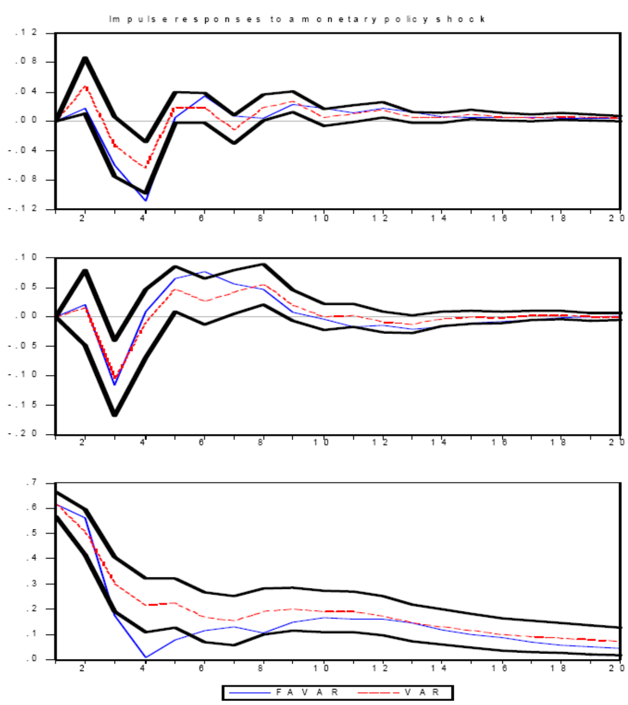

Figure 2: Responses of quarterly inflation, quarterly GDP growth and monetary policy rates to a monetary policy shock in the VAR and in the FAVAR(4)

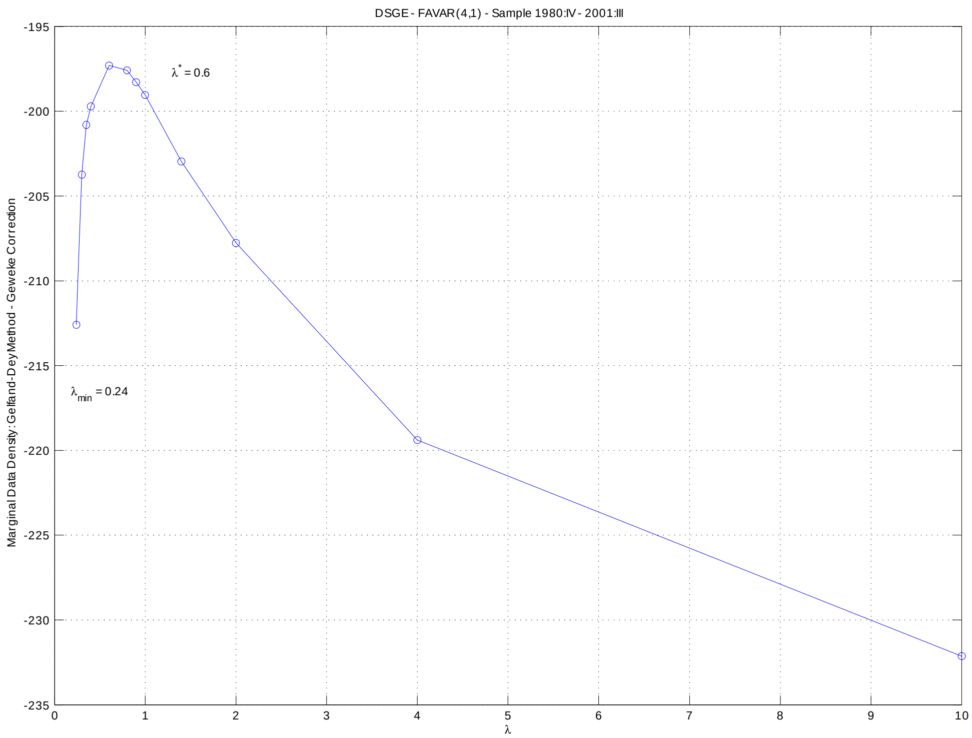

Figure 3: the optimal λ in a DSGE-FAVAR

2.9 Appendix A : The Sims Representation of our simple model

Del Negro and Schorfheide (2004) consider the following model:

$$\tilde{x}_t = E_t\tilde{x}_{t+1} - \frac{1}{\tau}(\tilde{R}_t - E_t\tilde{\pi}_{t+1}) + (1-\rho_G)\tilde{g}_t + \rho_z\frac{1}{\tau}\tilde{z}_t \tag{2.27}$$

$$\tilde{\pi}_t = \beta E_t\tilde{\pi}_{t+1} + \kappa\,(\tilde{x}_t - \tilde{g}_t) \tag{2.28}$$

$$\tilde{R}_t = \rho_R\tilde{R}_{t-1} + (1-\rho_R)(\psi_1\tilde{\pi}_t + \psi_2\tilde{x}_t) + \epsilon_{R,t} \tag{2.29}$$

$$\tilde{g}_t = \rho_g\tilde{g}_{t-1} + \epsilon_{g,t} \tag{2.30}$$

$$\tilde{z}_t = \rho_z\tilde{z}_{t-1} + \epsilon_{z,t} \tag{2.31}$$

The first step towards solution is to cast the model in the form of :

$$\Gamma_0\widetilde{\mathbf{Z}}_t = \Gamma_1\widetilde{\mathbf{Z}}_{t-1} + C + \Psi\epsilon_t + \Pi\eta_t \tag{2.32}$$

The results is achieved as follows:

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

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

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

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

2.10 Appendix B : The data used to extract factors

We describe data used to extract factors in the format adopted by Stock and Watson(2002):series number, long description, short description, transformation code and slow code (0. The transformation code are: 1 - no transformation; 2 - first difference; 3 - second difference; 4 - logarithm; 5 - first difference of logarithm and 6 - second difference of logarithm.

Date	Long Description	Short Desc	Transf cod	SlowCod
a0m052	Personal income (AR, bil. chain 2000 $)	PI	5	1
A0M051	Personal income less transfer payments (AR, bil. chain 2000 $)	PI less transfers	5	1
A0M224_	Real Consumption (AC) A0m224/gmdc	Consumption	5	1
A0M057	Manufacturing and trade sales (mil. Chain 1996 $)	M&T sales	5	1
A0M059	Sales of retail stores (mil. Chain 2000 $)	Retail sales	5	1
IPS10	INDUSTRIAL PRODUCTION INDEX - TOTAL INDEX	IP: total	5	1
IPS11	INDUSTRIAL PRODUCTION INDEX - PRODUCTS, TOTAL	IP: products	5	1
IPS299	INDUSTRIAL PRODUCTION INDEX - FINAL PRODUCTS	IP: final prod	5	1
IPS12	INDUSTRIAL PRODUCTION INDEX - CONSUMER GOODS	IP: cons gds	5	1
IPS13	INDUSTRIAL PRODUCTION INDEX - DURABLE CONSUMER GOODS	IP: cons dble	5	1
IPS18	INDUSTRIAL PRODUCTION INDEX - NONDURABLE CONSUMER GOODS	IP:cons nondbl	5	1
IPS25	INDUSTRIAL PRODUCTION INDEX - BUSINESS EQUIPMENT	IP:bus eqpt	5	1
IPS32	INDUSTRIAL PRODUCTION INDEX - MATERIALS	IP: matls	5	1
IPS34	INDUSTRIAL PRODUCTION INDEX - DURABLE GOODS MATERIALS	IP: dble mats	5	1
IPS38	INDUSTRIAL PRODUCTION INDEX - NONDURABLE GOODS MATERIALS	IP:nondble mats	5	1
IPS43	INDUSTRIAL PRODUCTION INDEX - MANUFACTURING (SIC)	IP: mfg	5	1
IPS307	INDUSTRIAL PRODUCTION INDEX - RESIDENTIAL UTILITIES	IP: res util	5	1
IPS306	INDUSTRIAL PRODUCTION INDEX - FUELS	IP: fuels	5	1
PMP	NAPM PRODUCTION INDEX (PERCENT)	NAPM prodn	1	1
A0m082	Capacity Utilization (Mfg)	Cap util	2	1
LHEL	INDEX OF HELP-WANTED ADVERTISING IN NEWSPAPERS (1967=100;S	Help wanted ind	2	1
LHELX	EMPLOYMENT: RATIO; HELP-WANTED ADS:NO. UNEMPLOYED CLF	Help wanted/em	2	1
LHEM	CIVILIAN LABOR FORCE: EMPLOYED, TOTAL (THOUS.,SA)	Emp CPS total	5	1
LHNAG	CIVILIAN LABOR FORCE: EMPLOYED, NONAGRIC.INDUSTRIES (THOUS	Emp CPS nona	5	1
LHUR	UNEMPLOYMENT RATE: ALL WORKERS, 16 YEARS & OVER (%,SA)	U: all	2	1
LHU680	UNEMPLOY.BY DURATION: AVERAGE(MEAN)DURATION IN WEEKS (SA	U: mean duratio	2	1
LHU5	UNEMPLOY.BY DURATION: PERSONS UNEMPL.LESS THAN 5 WKS (TH	U < 5 wks	5	1
LHU14	UNEMPLOY.BY DURATION: PERSONS UNEMPL.5 TO 14 WKS (THOUS.,	U 5-14 wks	5	1
LHU15	UNEMPLOY.BY DURATION: PERSONS UNEMPL.15 WKS + (THOUS.,SA)	U 15+ wks	5	1
LHU26	UNEMPLOY.BY DURATION: PERSONS UNEMPL.15 TO 26 WKS (THOUS.	U 15-26 wks	5	1
LHU27	UNEMPLOY.BY DURATION: PERSONS UNEMPL.27 WKS + (THOUS,SA)	U 27+ wks	5	1
A0M005	Average weekly initial claims, unemploy. insurance (thous.)	UI claims	5	1
CES002	EMPLOYEES ON NONFARM PAYROLLS - TOTAL PRIVATE	Emp: total	5	1
CES003	EMPLOYEES ON NONFARM PAYROLLS - GOODS-PRODUCING	Emp: gds prod	5	1
CES006	EMPLOYEES ON NONFARM PAYROLLS - MINING	Emp: mining	5	1
CES011	EMPLOYEES ON NONFARM PAYROLLS - CONSTRUCTION	Emp: const	5	1
CES015	EMPLOYEES ON NONFARM PAYROLLS - MANUFACTURING	Emp: mfg	5	1
CES017	EMPLOYEES ON NONFARM PAYROLLS - DURABLE GOODS	Emp: dble gds	5	1
CES033	EMPLOYEES ON NONFARM PAYROLLS - NONDURABLE GOODS	Emp: nondbles	5	1

CES046	EMPLOYEES ON NONFARM PAYROLLS - SERVICE-PROVIDING	Emp: services	5	1
CES048	EMPLOYEES ON NONFARM PAYROLLS - TRADE, TRANSPORTATION, A	Emp: TTU	5	1
CES049	EMPLOYEES ON NONFARM PAYROLLS - WHOLESALE TRADE	Emp: wholesale	5	1
CES053	EMPLOYEES ON NONFARM PAYROLLS - RETAIL TRADE	Emp: retail	5	1
CES088	EMPLOYEES ON NONFARM PAYROLLS - FINANCIAL ACTIVITIES	Emp: FIRE	5	1
CES140	EMPLOYEES ON NONFARM PAYROLLS - GOVERNMENT	Emp: Govt	5	1
A0M048	Employee hours in nonag. establishments (AR, bil. hours)	Emp-hrs nonag	5	1
CES151	AVERAGE WEEKLY HOURS OF PRODUCTION OR NONSUPERVISORY W	Avg hrs	1	1
CES155	AVERAGE WEEKLY HOURS OF PRODUCTION OR NONSUPERVISORY W	Overtime: mfg	2	1
aom001	Average weekly hours, mfg. (hours)	Avg hrs: mfg	1	1
PMEMP	NAPM EMPLOYMENT INDEX (PERCENT)	NAPM empl	1	1
HSFR	HOUSING STARTS:NONFARM(1947-58);TOTAL FARM&NONFARM(1959-)	HStarts: Total	4	0
HSNE	HOUSING STARTS:NORTHEAST (THOUS.U.)S.A.	HStarts: NE	4	0
HSMW	HOUSING STARTS:MIDWEST(THOUS.U.)S.A.	HStarts: MW	4	0
HSSOU	HOUSING STARTS:SOUTH (THOUS.U.)S.A.	HStarts: South	4	0
HSWST	HOUSING STARTS:WEST (THOUS.U.)S.A.	HStarts: West	4	0
HSBR	HOUSING AUTHORIZED: TOTAL NEW PRIV HOUSING UNITS (THOUS.,S	BP: total	4	0
HSBNE	HOUSES AUTHORIZED BY BUILD. PERMITS:NORTHEAST(THOU.U.)S.A	BP: NE	4	0
HSBMW	HOUSES AUTHORIZED BY BUILD. PERMITS:MIDWEST(THOU.U.)S.A.	BP: MW	4	0
HSBSOU	HOUSES AUTHORIZED BY BUILD. PERMITS:SOUTH(THOU.U.)S.A.	BP: South	4	0
HSBWST	HOUSES AUTHORIZED BY BUILD. PERMITS:WEST(THOU.U.)S.A.	BP: West	4	0
PMI	PURCHASING MANAGERS' INDEX (SA)	PMI	1	0
PMNO	NAPM NEW ORDERS INDEX (PERCENT)	NAPM new ord	1	0
PMDEL	NAPM VENDOR DELIVERIES INDEX (PERCENT)	NAPM vendor d	1	0
PMNV	NAPM INVENTORIES INDEX (PERCENT)	NAPM Invent	1	0
A0M008	Mfrs' new orders, consumer goods and materials (bil. chain 1982 $)	Orders: cons gds	5	0
A0M007	Mfrs' new orders, durable goods industries (bil. chain 2000 $)	Orders: dble gds	5	0
A0M027	Mfrs' new orders, nondefense capital goods (mil. chain 1982 $)	Orders: cap gds	5	0
A1M092	Mfrs' unfilled orders, durable goods indus. (bil. chain 2000 $)	Unf orders: dble	5	0
A0M070	Manufacturing and trade inventories (bil. chain 2000 $)	M&T invent	5	0
A0M077	Ratio, mfg. and trade inventories to sales (based on chain 2000 $)	M&T invent/sales	2	0
FM1	MONEY STOCK: M1(CURR,TRAV.CKS,DEM DEP,OTHER CK'ABLE DEP)(BIL$,SA)	M1	6	0
FM2	MONEY STOCK:M2(M1+O'NITE RPS,EURO$,G/P&B/D MMMFS&SAV&SM TIME DEP(B	M2	6	0
FM3	MONEY STOCK: M3(M2+LG TIME DEP,TERM RP'S&INST ONLY MMMFS)(BIL$,SA)	M3	6	0
FM2DQ	MONEY SUPPLY - M2 IN 1996 DOLLARS (BCI)	M2 (real)	5	0
FMFBA	MONETARY BASE, ADJ FOR RESERVE REQUIREMENT CHANGES(MIL$,SA)	MB	6	0
FMRRA	DEPOSITORY INST RESERVES:TOTAL,ADJ FOR RESERVE REQ CHGS(MIL$,SA)	Reserves tot	6	0
FMRNBA	DEPOSITORY INST RESERVES:NONBORROWED,ADJ RES REQ CHGS(MIL$,SA)	Reserves nonbor	6	0
FCLNQ	COMMERCIAL & INDUSTRIAL LOANS OUSTANDING IN 1996 DOLLARS (BCI)	C&I loans	6	0
FCLBMC	WKLY RP LG COM'L BANKS:NET CHANGE COM'L & INDUS LOANS(BIL$,SAAR)	C&I loans	1	0
CCINRV	CONSUMER CREDIT OUTSTANDING - NONREVOLVING(G19)	Cons credit	6	0

A0M095	Ratio, consumer installment credit to personal income (pct.)	Inst cred/PI	2	0
FSPCOM	S&P'S COMMON STOCK PRICE INDEX: COMPOSITE (1941-43=10)	S&P 500	5	0
FSPIN	S&P'S COMMON STOCK PRICE INDEX: INDUSTRIALS (1941-43=10)	S&P: indust	5	0
FSDXP	S&P'S COMPOSITE COMMON STOCK: DIVIDEND YIELD (% PER ANNUM)	S&P div yield	2	0
FSPXE	S&P'S COMPOSITE COMMON STOCK: PRICE-EARNINGS RATIO (%,NSA)	S&P PE ratio	5	0
FYFF	INTEREST RATE: FEDERAL FUNDS (EFFECTIVE) (% PER ANNUM,NSA)	FedFunds	2	0
CP90	Cmmercial Paper Rate (AC)	Commpaper	2	0
FYGM3	INTEREST RATE: U.S.TREASURY BILLS,SEC MKT,3-MO.(% PER ANN,NSA)	3 mo T-bill	2	0
FYGM6	INTEREST RATE: U.S.TREASURY BILLS,SEC MKT,6-MO.(% PER ANN,NSA)	6 mo T-bill	2	0
FYGT1	INTEREST RATE: U.S.TREASURY CONST MATURITIES,1-YR.(% PER ANN,NSA)	1 yr T-bond	2	0
FYGT5	INTEREST RATE: U.S.TREASURY CONST MATURITIES,5-YR.(% PER ANN,NSA)	5 yr T-bond	2	0
FYGT10	INTEREST RATE: U.S.TREASURY CONST MATURITIES,10-YR.(% PER ANN,NSA)	10 yr T-bond	2	0
FYAAAC	BOND YIELD: MOODY'S AAA CORPORATE (% PER ANNUM)	Aaabond	2	0
FYBAAC	BOND YIELD: MOODY'S BAA CORPORATE (% PER ANNUM)	Baa bond	2	0
scp90	cp90-fyff	CP-FF spread	1	0
sfygm3	fygm3-fyff	3 mo-FF spread	1	0
sFYGM6	fygm6-fyff	6 mo-FF spread	1	0
sFYGT1	fygt1-fyff	1 yr-FF spread	1	0
sFYGT5	fygt5-fyff	5 yr-FFspread	1	0
sFYGT10	fygt10-fyff	10yr-FF spread	1	0
sFYAAAC	fyaaac-fyff	Aaa-FF spread	1	0
sFYBAAC	fybaac-fyff	Baa-FF spread	1	0
EXRUS	UNITED STATES;EFFECTIVE EXCHANGE RATE(MERM)(INDEX NO.)	Ex rate: avg	5	0
EXRSW	FOREIGN EXCHANGE RATE: SWITZERLAND (SWISS FRANC PER U.S.$)	Ex rate: Switz	5	0
EXRJAN	FOREIGN EXCHANGE RATE: JAPAN (YEN PER U.S.$)	Ex rate: Japan	5	0
EXRUK	FOREIGN EXCHANGE RATE: UNITED KINGDOM (CENTS PER POUND)	Ex rate: UK	5	0
EXRCAN	FOREIGN EXCHANGE RATE: CANADA (CANADIAN $ PER U.S.$)	EX rate: Canada	5	0
PWFSA	PRODUCER PRICE INDEX: FINISHED GOODS (82=100,SA)	PPI: fin gds	6	0
PWFCSA	PRODUCER PRICE INDEX:FINISHED CONSUMER GOODS (82=100,SA)	PPI: cons gds	6	0
PWIMSA	PRODUCER PRICE INDEX:INTERMED MAT.SUPPLIES & COMPONENTS(82=100,SA)	PPI: int mat'ls	6	0
PWCMSA	PRODUCER PRICE INDEX:CRUDE MATERIALS (82=100,SA)	PPI: crude mat'ls	6	0
PSCCOM	SPOT MARKET PRICE INDEX:BLS & CRB: ALL COMMODITIES(1967=100)	Commod: spot pri	6	0
PSM99Q	INDEX OF SENSITIVE MATERIALS PRICES (1990=100)(BCI-99A)	Sens mat'ls price	6	0
PMCP	NAPM COMMODITY PRICES INDEX (PERCENT)	NAPM com price	1	0
PUNEW	CPI-U: ALL ITEMS (82-84=100,SA)	CPI-U: all	6	1
PU83	CPI-U: APPAREL & UPKEEP (82-84=100,SA)	CPI-U: apparel	6	1
PU84	CPI-U: TRANSPORTATION (82-84=100,SA)	CPI-U: transp	6	1
PU85	CPI-U: MEDICAL CARE (82-84=100,SA)	CPI-U: medical	6	1
PUC	CPI-U: COMMODITIES (82-84=100,SA)	CPI-U: comm.	6	1
PUCD	CPI-U: DURABLES (82-84=100,SA)	CPI-U: dbles	6	1
PUS	CPI-U: SERVICES (82-84=100,SA)	CPI-U: services	6	1
PUXF	CPI-U: ALL ITEMS LESS FOOD (82-84=100,SA)	CPI-U: ex food	6	1
PUXHS	CPI-U: ALL ITEMS LESS SHELTER (82-84=100,SA)	CPI-U: ex shelter	6	1
PUXM	CPI-U: ALL ITEMS LESS MEDICAL CARE (82-84=100,SA)	CPI-U: ex med	6	1
GMDC	PCE,IMPL PR DEFL:PCE (1987=100)	PCE defl	6	1
GMDCD	PCE,IMPL PR DEFL:PCE; DURABLES (1987=100)	PCE defl: dlbes	6	1
GMDCN	PCE,IMPL PR DEFL:PCE; NONDURABLES (1996=100)	PCE defl: nondble	6	1
GMDCS	PCE,IMPL PR DEFL:PCE; SERVICES (1987=100)	PCE defl: services	6	1
CES275	AVERAGE HOURLY EARNINGS OF PRODUCTION OR NONSUPERVISORY WORKER	AHE: goods	6	1
CES277	AVERAGE HOURLY EARNINGS OF PRODUCTION OR NONSUPERVISORY WORKER	AHE: const	6	1
CES278	AVERAGE HOURLY EARNINGS OF PRODUCTION OR NONSUPERVISORY WORKER	AHE: mfg	6	1

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

Here we provide the full derivation of the results reported in Section 3 on the DS approach to obtain draws from the posterior distribution of (Φ, Σ_u, θ). The analysis will be conditional to a value for λ which establishes the relevance of the information between the VAR and DSGE in order to estimate the structural parameter θ. We can think of λ as generating a particular model which can support, with a certain degree, the observed data: the marginal data density represents such a measure of goodness and it would help us to discriminate among different models (i.e. different λ).

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

2.11.1 The Bayesian Approach

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

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

The VAR model for the data is

$$Y_t = \Phi X_t + E_t, \tag{2.33}$$

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

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

$$\tilde{Y}_t = \Lambda_0(\theta) + \Lambda_1(\theta)\tilde{Z}_t + V_t, \tag{2.34}$$

$$\tilde{Z}_t = T(\theta)\tilde{Z}_{t-1} + R(\theta)U_t, \tag{2.35}$$

which groups the policy function from the RE equilibrium and the mapping between observables, $\tilde{Y}_t$, and simulated data, $\tilde{Z}_t$. The vector $\tilde{Y}_t$ can be computed by simulation methods with respect to (2.34) and (2.35) or analytically since the DSGE model is stationary.

Given the pair of simulated data $\left(\tilde{Y}_t, \tilde{X}_t\right)$[15] we can write a similar specification as in (2.33)

$$\tilde{Y}_t = \Phi\tilde{X}_t + E_t, \tag{2.36}$$

that indirectly imposes restrictions on Φ driven from the theoretical model; to derive the DSGE-based prior we will construct the likelihood function of the process in (2.36).

2.11.2 Compute DSGE Moments

Given the state-space representation in (2.34) and (2.35), the unconditional variance for $\tilde{Y}_t$ and $\tilde{Z}_t$ are

$$\Sigma_{z,z} = T\Sigma_{z,z}T' + R\Sigma_{u,u}R' \tag{2.37}$$

$$\Sigma_{y,y} = \Lambda_0\Lambda_0' + \Lambda_1\Sigma_{z,z}\Lambda_1' + \Sigma_{v,v} + \Lambda_1 R\Sigma_{u,v} + \Sigma_{u,v}' R'\Lambda_1' \tag{2.38}$$

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

$$\Sigma_{z,z}(k) = T^k\Sigma_{z,z}(k-1) \tag{2.39}$$

$$\Sigma_{y,y}(k) = \Lambda_0\Lambda_0' + \Lambda_1\Sigma_{z,z}(k)\Lambda_1' + \Lambda_1\left(T^k\right)R\Sigma_{u,v}. \tag{2.40}$$

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

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

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

The likelihood function for the artificial data reads

$$\mathcal{L}\left(\tilde{Y};\Phi,\Sigma_e\right)=(2\pi)^{-mT/2}\left|\Sigma_e\right|^{-T/2}\exp\left(-\frac{1}{2}tr\left(\left(\left(\Phi-\tilde{\Phi}\right)\left(\tilde{X}'\tilde{X}\right)\left(\Phi-\tilde{\Phi}\right)+\tilde{S}\right)\Sigma_e^{-1}\right)\right), \tag{2.41}$$

where the sufficient statistics are,

$$\tilde{\Phi} = \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{Y} \tag{2.42}$$

$$\tilde{S} = \tilde{Y}'\tilde{Y}-\tilde{Y}'\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{Y} \tag{2.43}$$

which can be also specified in terms of population moments

$$\tilde{\Phi} = \Sigma_{x,x}^{-1}\Sigma_{x,y} \tag{2.44}$$

$$\tilde{S} = \Sigma_{y,y}-\Sigma_{x,y}'\Sigma_{x,x}^{-1}\Sigma_{x,y} \tag{2.45}$$

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

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

$$\pi_0=\left|\Sigma_e\right|^{-\frac{m+1}{2}}. \tag{2.46}$$

By combining (2.41) and (2.46) we get the kernel of the distribution

$$\pi_1\propto\mathcal{L}\left(\tilde{Y}\mid\Phi,\Sigma_e\right)\times\pi_0, \tag{2.47}$$

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

$$P_{\tilde{Y}}\left(\tilde{Y}\mid\theta\right)=(2\pi)^{-m\tilde{v}/2}\times\left|\tilde{S}\right|^{-\frac{\tilde{v}}{2}}\times\left|\tilde{H}\right|^{-\frac{m}{2}}\times\left[2^{m\tilde{v}/2}\times\pi^{m(m-1)/4}\times\Gamma_m\left(\tilde{v}\right)\right], \tag{2.48}$$

which is needed to have the DSGE-based prior distribution

$$\begin{aligned}
\pi_1\left(\Phi, \Sigma_e \mid \tilde{Y}, \theta\right) &= \frac{\mathcal{L}\left(\tilde{Y} \mid \Phi, \Sigma_e, \theta\right) \times \pi_0}{P_{\tilde{Y}}\left(\tilde{Y} \mid \theta\right)} \\
&= \frac{(2\pi)^{-mT^*/2}}{(2\pi)^{-m\tilde{v}/2}} \times \frac{\left|\tilde{S}\right|^{\tilde{v}/2} \times \left|\tilde{H}\right|^{m/2} \times \left|\Sigma_e\right|^{-(T^*+m+1)/2}}{2^{m\tilde{v}/2} \times \pi^{m(m-1)/4} \times \Gamma_m\left(\tilde{v}\right)} \times \\
&\quad \exp\left[-\frac{1}{2} tr\left(\tilde{S}\Sigma_e^{-1}\right)\right] \times \exp\left[-\frac{1}{2} tr\left(\left(\Phi - \tilde{\Phi}\right)' \left(\Sigma_{x,x}\right) \left(\Phi - \tilde{\Phi}\right) \Sigma_e^{-1}\right)\right],
\end{aligned} \tag{2.49}$$

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

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

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

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

$$P_{\Phi}\left(\Phi, \Sigma_e \mid Y, \theta\right) \propto \mathcal{L}\left(Y \mid \Phi, \Sigma_e\right) \times \pi_1\left(\Phi, \Sigma_e \mid \tilde{Y}, \theta\right), \tag{2.50}$$

which can be integrated to obtain the marginal data density[16]

$$P_Y\left(Y \mid \theta\right) = (2\pi)^{-Tm/2} \times \frac{\left|\tilde{S}\right|^{\tilde{v}/2}}{\left|\bar{S}\right|^{\bar{v}/2}} \frac{\left|\tilde{H}\right|^{m/2}}{\left|\bar{H}\right|^{m/2}} \times \frac{\Gamma_m\left(\bar{v}\right)}{\Gamma_m\left(\tilde{v}\right)} \times 2^{m(\hat{v}+k)/2} \tag{2.51}$$

[16] where

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

The proper posterior reads

$$P_\Phi\left(\Phi, \Sigma_e \mid Y, \theta\right) = (2\pi)^{-mk/2} \times |\Sigma_e|^{-k/2} \times \exp\left[-\frac{1}{2}tr\left(\left(\Phi - \bar{\Phi}\right)' \bar{H}\left(\Phi - \bar{\Phi}\right)\Sigma_e^{-1}\right)\right]\ldots$$
$$\times\frac{\left|\bar{S}\right|^{\bar{v}/2}\left|\bar{H}\right|^{m/2} \times |\Sigma_e|^{-(\hat{v}+T^*+m+1)/2}}{2^{m\bar{v}/2} \times \pi^{m(m-1)/4}\Gamma_m\left(\bar{v}\right)} \times \exp\left[-\frac{1}{2}tr\left(\bar{S}\Sigma_e^{-1}\right)\right] \quad (2.52)$$

or equivalently

$$p\left(\Phi \mid \Sigma_e; Y, X\right) = N\left(\bar{\Phi}, \Sigma_e \otimes \bar{H}^{-1}\right) \quad (2.53)$$
$$p\left(\Sigma_e \mid Y, X\right) = IW\left(\bar{S}, \bar{v}\right) \quad (2.54)$$

where the posterior estimates are as follows

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

2.11.5 Metropolis-Hasting Algorithm

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

$$P\left(\Phi, \Sigma, \theta \mid Y\right) = P_\Phi\left(\Phi, \Sigma \mid Y, \theta\right) \times P_\theta\left(\theta \mid Y\right). \quad (2.55)$$

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

$$P_\theta\left(\theta \mid Y\right) \propto \mathbb{K}_\theta\left(\theta \mid Y\right) = P_Y\left(Y \mid \theta\right) \times \pi_2\left(\theta\right) \quad (2.56)$$

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

$$\tilde{\theta} = \theta_{s-1} + (\Sigma_H(\theta^*))^{-1/2} N(0, I), \tag{2.57}$$

we construct an acceptance probability threshold

$$\alpha\left(\tilde{\theta}, \theta_{s-1}\right) = \min\left(1, \frac{\mathbb{K}_\theta\left(\tilde{\theta} \mid Y\right)}{\mathbb{K}_\theta(\theta_{s-1} \mid Y)}\right). \tag{2.58}$$

If $\alpha\left(\tilde{\theta}, \theta_{s-1}\right)$ is higher than a certain probability (varying for each draw) we accept the draw as coming from the posterior distribution $P_\theta(\theta \mid Y)$ and update the Markov chain $\theta_s = \tilde{\theta}$, otherwise we discard $\tilde{\theta}$ and draw another candidate from (2.57).

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

2.11.6 Gelfand-Dey Method for $P(Y)$

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

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

2.11.7 A FAVAR Analysis of the Simple DSGE Model

In this case the benchmark specification for the unrestricted dynamics of the variables included in the theoretical model becomes the following:

$$Y_t = B_0 X_t + B_1 F_t + E_t \tag{2.59}$$

where $\mathbf{Y}_t = (\Delta \ln x_t, \Delta \ln P_t, \ln R_t)$, $X_t = [1, \mathbf{Y}_{t-1}, ...\mathbf{Y}_{t-p}]$, $F_t = \left[f_t', f_{t-1}', \ldots f_{t-q}'\right]'$ groups q lags of the three factors $f_t = [f_{1,t}, f_{2,t}, f_{3,t}]'$ extracted and interpreted as in Bernanke, Boivin and Eliasz (2005), E_t is the three-variate vector of innovations. System (2.59) can be re-written in a more compact form as follows:

$$Y_t = BW_t + E_t \tag{2.60}$$

where $B = [B_0, B_1]$ is of dimension $m \times (1 + mp + rq)$ and $W_t = [X_t', F_t']'$.

Prior distribution

The full prior on the coefficients in (2.60) is derived by recalling the moments from the DSGE model as we did in Section 3 and by working out a prior for the factors coefficients which is centered at zero with a variance-covariance matrix set by the second moments matrix of the factors. Given that factors do not enter in the DSGE model, we can draw dummy observations from the theoretical model for the endogenous variables, $\left(\tilde{Y}_t, \tilde{X}_t\right)$,[17] without considering the effect from $\tilde{F}_t$. At the same time we can derive dummy observations to set the prior on the coefficients of the factors, $\tilde{F}_t$, by using a training sample on the full FAVAR. The set of dummy observations $\left(\tilde{Y}_t, \tilde{X}_t, \tilde{F}_t\right)$ can be used to derive the full prior distribution over the coefficients which reads

$$\begin{bmatrix} B_0 \\ B_1 \end{bmatrix} \mid \Sigma_e \sim N\left(\begin{bmatrix} \tilde{B}_0 \\ 0 \end{bmatrix}, \Sigma_e \otimes \begin{bmatrix} \left(\tilde{X}_t \tilde{X}_t'\right)^{-1} & 0 \\ 0 & \left(\tilde{F}_t \tilde{F}_t'\right)^{-1} \end{bmatrix}\right) \tag{2.61}$$

[17] As in the DSGE-VAR, their population counterparts are used

where $\tilde{B}_0 = \left(\tilde{X}_t \tilde{X}_t'\right)^{-1} \tilde{X}_t \tilde{Y}_t'$. The cross term restriction, $\left(\tilde{X}_t \tilde{F}_t'\right)$, is also set to zero because, in constructing our prior, we are considering the case in which factors don't have any influence on the set of endogenous variables in our DSGE model. We spell out all these details in Appendix C. As far as the prior distribution for the structural parameters is concerned, we maintain the same independence assumption as we did in Section 3; we also consider the same shape and parameterization.

Posterior distribution

Given our description of the prior distribution and the likelihood function we can proceed with the illustration of the computation of the posterior distribution. A new feature of the analysis at this stage has to do with the contribution of the factors in shaping inference. The DSGE model itself does not directly depend on factors, but its estimates account for the larger information set as it appears from the following decomposition

$$P\left(\Phi, \Sigma_e, \theta \mid Y, F\right) = P_{\Phi}\left(\Phi, \Sigma_e \mid \theta, Y, F\right) \times P_{\theta}\left(\theta \mid Y, F\right), \tag{2.62}$$

where the posterior for θ, $P_{\theta}\left(\theta \mid Y, F\right)$, makes clear the dependence on the factors.

Chapter 3

Assessing the Potential of DSGE Model Evaluation in a Bayesian Framework

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 representation 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.

Keywords: Bayesian Analysis, DSGE Models, Vector Autoregressions

JEL Classification: C11, C15, C32

3.1 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.

The crucial point is the possibility to find an econometric representation for a theoretical model. Since the early 1980s, two distinct approaches have emerged.

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. Despite of the popularity in the use of VAR models for the data description and forecasting exercises, they are subject to Lucas critique (Lucas, 1976), to multicollinearity problems and they can fail to take to account of non-linearities in the economy.

The second approach consists of the use of DSGE model to represent the economy, but using a theoretical model, there could be problems with the calibration used in the estimation of the model. The calibrated DSGE models are typically too stylized to be taken directly to the data and often yield fragile results (Smets and Wouters, 2003 and Ireland, 2004).

Recently, Doan et al. (1984), Ingram and Whiteman (1994), DeJong et al. (1996 and 2000), Schorfheide (2000) have used a Bayesian framework to estimate and evaluate DSGE Models. In the "New macroeconometrics" literature, the principle underling a Bayesian analysis of DSGE Models is to combine prior and likelihood functions in order to obtain posterior distribution of the interested variables.

A prominent new procedure applied to estimate DSGE models by using VAR approach and Bayesian econometric is presented by two influential papers, Del Negro and Schorfheide (2004) and Del Negro, Schorfheide, Smets and Wouters (2007a). This methodology helps economists to choose the best model to represent real data among a theoretical framework from the economic literature, a statistical representation from the data and a combination between the two 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 Ingram and Whiteman (1994) and 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 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 [1] , 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

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

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 and a Bayesian VAR implied by the econometric framework of the economic model. The theoretical 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.

For example, in Liu et al. (2008), DSGE-VAR is used to forecast South African Economy, in Adjemian et al. (2008), DSGE-VAR is used in order to compare different optimal monetary policy and in Adolfson et al. (2008) and Lees et al. (2007), the hybrid model, DSGE-VAR, 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[2]. For example, the comments provided by two papers could be considered useful in order to understand better how DSGE-VAR is implemented. The first, 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

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

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 (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 by the lags used in the VAR representation[3]).

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 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 under the null hypothesis and alternative hypothesis. Under the null hypothesis, the Data Generating Process (DGP) come from the estimated model and under alternative hypothesis these artificial data are not explained by the estimated model.

In the first MonteCarlo experiment, the DGP is the forward-looking model which is the candidate model in order to explain the data. 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

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

parameter sets and not from the forward-looking model. Under the null hypothesis, DGP come from the candidate model, but under alternative hypothesis, the DGP come from the backward-looking model.

The purpose of these experiments is to understand if the DSGE-VAR has the advantage to recognize if the DGP is the model used to generate the 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.

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 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, 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 4.

3.2 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 the 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 and the characteristics of the candidate economic model to explain data in the estimation, used in this procedure to generate the 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.

3.2.1 DSGE-VAR Approach: a general assessment

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):

$$Y = X\Phi + U \tag{3.1}$$

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:

$$L(Y|\Phi, \Sigma_u) \tag{3.2}$$

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):

$$\begin{aligned} x_{t+1} &= A(\theta)x_t + B(\theta)w_t \\ y_t &= C(\theta)x_t + D(\theta)w_t \end{aligned} \tag{3.3}$$

where w_t is an $k \times 1$ vector of structural shocks satisfying $E[w_t] = 0, E[w_t w_t'] = I$ and $E[w_t w_{t-j}] = 0$ 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:

$$y_t = \sum_{j=1}^{\infty} C\left(A - BD^{-1}C\right)^{j-1} BD^{-1} y_{t-j} + Dw_t \tag{3.4}$$

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 :

$$\begin{aligned} \Phi^*(\theta) &= \Gamma^{*-1}_{xx}(\theta)\,\Gamma^*_{xy}(\theta) \qquad (3.5) \\ \Sigma^*(\theta) &= \Gamma^*_{yy}(\theta) - \Gamma^*_{yx}(\theta)\,\Gamma^{*-1}_{xx}(\theta)\,\Gamma^*_{xy}(\theta) \end{aligned}$$

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:

$$Y^* = X^*\Phi(\theta) + U^* \qquad (3.6)$$

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.

Following the Bayesian econometrics literature suggesting the use of dummy observation priors[4] to impose a prior distribution on the set of coefficients, in the Del Negro and Schorfheide procedure, these dummy observations 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

[4] 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.

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.

$$P(\Phi, \Sigma_u, \theta, \lambda) = P\left(\Phi, \Sigma_u | \theta, \lambda\right) P\left(\theta\right) \tag{3.7}$$

In the previous formula of a hierarchical prior, $P\left(\theta\right)$ represents the prior from the model; $P\left(\Phi, \Sigma_u | \theta, \lambda\right)$ is composed by the 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 by the econometrician. When λ approaches ∞, the prior converges to a single spike over $\Phi(\theta)$ and $\Sigma_u\left(\theta\right)$ 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.

Posterior Distribution and Marginal Likelihood Function

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

$$P(\Phi, \Sigma_u, \theta | Y, \lambda) \propto L(Y | \Phi, \Sigma_u) P\left(\Phi, \Sigma_u | \theta, \lambda\right) P\left(\theta\right) \tag{3.8}$$

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 the vector 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:

$$L(Y,\lambda) = \int_{\Phi,\Sigma_u} L(Y|\Phi,\Sigma_u) P(\Phi,\Sigma_u|\theta,\lambda)\, d(\Phi,\Sigma_u) \tag{3.9}$$

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[5].

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.

3.2.2 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:

$$y_t = \phi y_{t-1} + u_t \qquad u_t \sim N(0,1) \tag{3.10}$$

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.

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

[5] 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).

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

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:

$$\ln p(Y|\lambda, \phi^*) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\widetilde{\sigma}^2(\lambda, \phi^*) - \frac{1}{2}c(\lambda, \phi^*) \tag{3.11}$$

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

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

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

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

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

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.

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

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 has as the minimum value a 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, $\widehat{\lambda}$, 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 log likelihood are given by the differences in the penalty terms. For large values of λ, penalty differentials become less important and 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 does not depend directly on the variables of the AR representation and on the the priors of the model, but only on the autocovariance and on the hyperparameter λ.

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[6].

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

[6] 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.

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 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, $\widehat{\lambda}$. The penalty term of the marginal likelihood function in the DSGE-VAR procedure can be compared to the penalty term in the usual Information Criteria.

3.2.3 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):

$$\int_{\Theta} \exp\left\{nh\left(\theta\right)\right\} d\theta = \exp\left\{nh\left(\theta\right)\right\} \left(2\pi\right)^{p/2} n^{-p/2} \left|H^{-1}\left(\widehat{\theta}\right)\right| + O(n^{-1}) \tag{3.12}$$

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:

$$B_{12}^{\pi} \simeq \frac{L_{1,n}\left(\widehat{\theta}_{1,n}\right)}{L_{2,n}\left(\widehat{\theta}_{2,n}\right)} \left| \frac{H^{-1}\left(\widehat{\theta}_{1,n}\right)}{H^{-1}\left(\widehat{\theta}_{2,n}\right)} \right|^{1/2} \left(\frac{n}{2\pi}\right)^{(p_2-p_1)/2} \tag{3.13}$$

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 :

$$\frac{L_{1,n}\left(\widehat{\theta}_{1,n}\right)}{L_{2,n}\left(\widehat{\theta}_{2,n}\right)} \tag{3.14}$$

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.

The Bayes Factor could be transformed as:

$$\log(B_{12}^{\pi}) \simeq \log \omega_n + \frac{p_2 - p_1}{2}\log(n) + K(\widehat{\theta}_{1,n}, \widehat{\theta}_{2,n}) \tag{3.15}$$

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

$$\omega_n = \frac{L_{1,n}\left(\widehat{\theta}_{1,n}\right)}{L_{2,n}\left(\widehat{\theta}_{2,n}\right)} \tag{3.16}$$

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

$$S = -\log \omega_n - \frac{p_2 - p_1}{2} \log(n) \tag{3.17}$$

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 y 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[7].

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

[7] 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.

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.

However, in this paper the classical Information Criteria, AIC, SIC and HQ (Hannah-Quinn) are used in a model comparison, ignoring the possible problems concerning the priors weight. It is possible to consider the overall DSGE-VAR, by using as a new "criterion" the maximization of the marginal data density, in order to check what is the lag which allows the econometrician the maximum value of the marginal data density, combining both the actual data and the artificial data. In both cases, the main idea is to stress out how the marginal likelihood depends on the number of parameters estimated.

3.2.4 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[8].

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

$$L(Y|\Phi, \Sigma_u) \tag{3.18}$$

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

[8] 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.

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 input 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:

$$\widetilde{x_t} = E_t[\widetilde{x_{t+1}}] - \frac{1}{\tau}(\widetilde{R_t} - E_t[\widetilde{\pi_{t+1}}]) + (1-\rho_G)\widetilde{g_t} + \rho_Z \frac{1}{\tau}\widetilde{z_t} \tag{3.19}$$

$$\widetilde{\pi_t} = \beta E_t[\widetilde{\pi_{t+1}}] + \kappa[\widetilde{x_t} - \widetilde{g_t}] \tag{3.20}$$

$$\widetilde{R_t} = \rho_R \widetilde{R_{t-1}} + (1-\rho_R)(\psi_1 \widetilde{\pi_t} + \psi_2 \widetilde{x_t}) + \epsilon_{R,t} \tag{3.21}$$

$$\widetilde{g_t} = \rho_g \widetilde{g_{t-1}} + \epsilon_{g,t} \tag{3.22}$$

$$\widetilde{z_t} = \rho_z \widetilde{z_{t-1}} + \epsilon_{z,t} \tag{3.23}$$

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.

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.

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 :

$$\Gamma_0 \widetilde{\mathbf{Z}}_t = \Gamma_1 \widetilde{\mathbf{Z}}_{t-1} + C + \Psi \epsilon_t + \Pi \eta_t \tag{3.24}$$

$t = 1, ..., T$ 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:

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

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

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

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

$$\widetilde{\mathbf{Z}}_t = \mathbf{T}(\theta)\,\widetilde{\mathbf{Z}}_{t-1} + \mathbf{R}(\theta)\,\varepsilon_t \tag{3.25}$$

$$\theta = [\kappa, \tau, \psi_1, \psi_2, \rho_R, \rho_g, \rho_Z, \sigma_R, \sigma_g, \sigma_Z]' \tag{3.26}$$

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:

$$\mathbf{Z}_t = W(\theta)\widetilde{\mathbf{Z}}_t + D(\theta) + \upsilon_t \tag{3.27}$$

where

$$\mathbf{Z}_t = \begin{bmatrix} \Delta \ln x_t \\ \Delta \ln P_t \\ \ln R_t^a \end{bmatrix} \tag{3.28}$$

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

$$\begin{aligned} \Delta \ln x_t &= \ln \gamma + \Delta \widetilde{x}_t + \widetilde{z}_t \\ \Delta \ln P_t &= \ln \pi^* + \widetilde{\pi}_t \\ \ln R_t^a &= 4[(\ln r^* + \ln \pi^*) + \widetilde{R}_t] \end{aligned} \tag{3.29}$$

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

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

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_t^*, X_t^* 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:

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

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 theoretical model chosen, 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.

3.3 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) and the estimated model are the same and under alternative hypothesis these artificial data do not come from the estimated model, but they come from an alternative model.

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 model used to generate the 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 come from a different model, 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.

3.3.1 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 the dummy observation priors and it is the candidate to explain the actual data.

In order to generate artificial data[9], 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

[9] 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.

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, $\widehat{\lambda}$, 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 [10] for each case of VAR representation, considering the different lags, are included in the lambda grid.[11]. Artificial data are generated by a MonteCarlo experiment with 100 replications[12].

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):

[10] 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).

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

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.

[12] 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.

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[13].

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 more 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 DGP as a VAR(1), the maximization of the marginal likelihood in order to assess the right number of lags in VAR representation 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.

In the next table, the usual Information Criteria [14] 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 the economic model is able to to explain the data over the pure statistical representation.

[13] 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 is considered.

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

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.

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.

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.

3.3.2 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:

$$\pi_t = \alpha_{\pi_1}\pi_{t-1} + \alpha_{\pi_2}\pi_{t-2} + \alpha_{\pi_3}\pi_{t-3} + \alpha_{\pi_4}\pi_{t-4} + \alpha_y y_{t-1} + \varepsilon_t^{\pi} \tag{3.30}$$

$$y_t = \beta_{y_1} y_{t-1} + \beta_{y_2} y_{t-2} + \beta_r \sum_{j=1}^{4} \frac{1}{4} (i - \pi)_{t-j} + \varepsilon_t^{y} \tag{3.31}$$

$$i_t = \gamma i_{t-1} + \varepsilon_t^{i} \tag{3.32}$$

The AS equation, (3.30), 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).

The AD equation, (3.31), 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 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, (3.32), 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, DGP comes from two 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. 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 real data on 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. The aim of these experiments is to generate artificial data from a model which is not the model used to generate the dummy observation priors and the candidate to explain the data, understanding if this model can explain the actual data. The possibility to recognize not only the misspecification of DGP but also the misspecification led by the lag-length can be considered as a crucial point. These MonteCarlo experiments are useful to check if the null hypothesis in the MonteCarlo experiment is not rejected, i.e. the forward-looking model can describe the backward-looking data if more lags are used in VAR representation.

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

$$X_t = AX_{t-1} + \upsilon_t \tag{3.33}$$

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[15]. The artificial data are generated by a MonteCarlo experiment with 100 replications[16].

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

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 DSGE-VAR composition. The expected result is to get the optimal lambda equal to the minimum lambda. In every DSGE-VAR exercise, the

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

[16] 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.

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 the 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 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, US 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). Moreover, the priors used to generate the data from the forward-looking model refer to the period of 1981-2001, used by Del Negro and Schorfheide (2004) in their application.

As before, by using MonteCarlo experiment, the artificial data with 80 observations are generated 100 times from the backward-looking model[17]. The artificial data are generated by using 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.

[17] 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 Data Generating Process is, as before, a VAR with 3 lags. 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 [18] 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 artificial data and the ratio increases after the first lag in 69% of cases. This exercise evidences, beyond the lags misspecification issue, that there is another problem concerning the misspecification, the artificial data comes from another model. However, this empirical analysis is useful to stress out that the DSGE-VAR collapses to a VAR in case of the misspecification of the economic model, but adding lags the economic model becomes relevant. In this case, it could be useful to consider a specific Information Criteria which takes into account the economic model, for example aforementioned Deviance Information Criterion (DIC) could be useful.

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

Forecasting

The forecasting performance concludes these MonteCarlo experiments. In this first table, the forecasting evaluation in MonteCarlo experiment is presented in case in which 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 in case in which 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.

3.3.3 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 data are representable by using a VAR with one lag and the best model is the statistical representation.

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 the 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 and it happens if the number of lags is increased. In this case, the misspecification leads to a more contribution from 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 crucial 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, in case 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.

3.3.4 Empirical Results in the Real World

The MonteCarlo experiments show the importance of the Data Generating Process in the DSGE-VAR procedure, but it is also important the correct choice of the number of lags. In this section, there is an empirical analysis in the real world, in order to understand what happens to the final result represented by the hybrid model by changing the sample size and moreover the number of lags. The purpose of these exercises is to check the marginal likelihood and its properties when there is a change in the minimum lambda used in the lambda grid.

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 based on the results obtained by Del Negro and Schorfheide (2004) deals with the sample size. In Del Negro and Schorfheide, VAR representation spans from the first quarter of 1981 to the last quarter of 2001, i.e. the sample size considers 80 observations. In this empirical analysis, the lambda grid used is different from the one used by Del Negro and Schorfheide (2004) in their implementation. This new set of weights is more precise around the possible optimal lambda (in Del Negro and Schorfheide replication, $\widehat{\lambda} = 0.6$) $\Lambda = \{0, 0.2, 0.4, 0.5, 0.6, 0.7, 1, 1.4, 1.8, 10\}$. 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 0.3. The new lambda grid is $\Lambda = \{0, 0.10, 0.15, 0.2, 0.4, 0.5, 0.6, 0.7, 1, 1.4, 1.8, 10\}$[19]. The optimal λ is 0.3, instead of 0.6.

The following table represents Del Negro and Schorfheide (2004) results using 25, 000 replications (as in their paper) in Metropolis Hastings.

[19] 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.

TABLE 12. VAR(4) Sample 1981-2001		
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

In the next table, the same exercise is replicated by using a larger sample and 25,000 replications in Metropolis Hastings.

TABLE 13. VAR(4) Sample 1961-2001		
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 in 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 $\widehat{\lambda}$[20] and its marginal data density, considering lags from 1 to 8, in case of 100,000 replications in Metropolis

[20] In this case the lambda grid takes into consideration the minimun lambda for each lag, $\Lambda = \{0, 0.09, 0.10, 0.13, 0.17, 0.2, 0.24, 0.25, 0.3, 0.31, 0.35, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$

Algorithm.

TABLE 14. Maximizing Marginal Data Density 100,000 replications

	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 in this two cases, the marginal data density is maximized with only one lag.

It is possible to consider the usual information criteria[21], the three real time series of US economy could be represented in a VAR framework by taking into consideration a more parsimonious model. Schwarz and Hannan-Quinn criteria suggest the use of only one lag, 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[22].

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

The next table shows the marginal likelihood 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.

[21] 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.

[22] 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			
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[23].

In the following table, the ratios calculated in case of VAR with four lags are compared with the ratios computed in case of a more parsimonious VAR.

[23] 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.
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.

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.

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. And, the large sample considers 160 quarters from 1961:01 to 2001:04. The $\widehat{\lambda}$ has been found for this 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 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.

3.4 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, considering the forecasting performance.

In this sense, it is important to truncate the VARMA representation for the theoretical model, 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.

Bibliography

[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): pp1-60.

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

[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.

3.5 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						