On the Statistical Identification of DSGE models*

Agostino Consolo, Carlo A. Favero, Alessia Paccagnini†

June 29, 2007

Abstract

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 Introduction

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[1]. As a consequence of this development, methods for diagnosing the fit of these models are being proposed and implemented. This article illustrates how the concept of statistical identification, originally

*We thank Marco Del Negro for having kindly provided us with the Matlab programs used in Del Negro-Schorfheide(2004).

†Consolo: IGIER(Università Bocconi). Favero: IGIER (Universita' Bocconi) and CEPR. Paccagnini, IGIER (Università Bocconi).

[1] 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.Killian.

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 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 illustrate the potential importance of the lack of statistical identification by showing that statistical identification can be achieved by using a Factor Augmented VAR (FAVAR), and by comparing 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 Statistical Identification of Cowles Commission and DSGE models

Spanos(1990) illustrates the importance of statistical identification for Cowles Commission models by considering the case of a simple demand and supply model on the market for commercial loans discussed in Maddala(1988). In the Cowles Commission tradition 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:

$$\begin{bmatrix} 1 & \gamma_{12} \\ 1 & \gamma_{22} \end{bmatrix} \begin{bmatrix} q_t \\ r_t \end{bmatrix} = \begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{13} & 0 & 0 \\ \delta_{21} & 0 & 0 & \delta_{24} & \delta_{25} \end{bmatrix} \begin{bmatrix} 1 \\ br_t \\ x_t \\ d_t \\ i_t \end{bmatrix} + \begin{bmatrix} u_t^d \\ u_t^s \end{bmatrix}$$

$$\begin{aligned} q_t^d &= \underset{(74.31)}{-210.43} - \underset{(1.60)}{20.2}\, r_t + \underset{(2.84)}{40.77} br_t + \underset{(0.45)}{2.34}\, x_t + \hat{u}_t^d \\ q_t^s &= \underset{(13.96)}{-87.94} + \underset{(1.89)}{6.09}\, r_t - \underset{(2.27)}{7.08}\, i_t + \underset{(0.008)}{0.334} d_t + \hat{u}_t^s \\ q_t^d &= q_t^s = q_t \\ \xi_1(1) &= 28.106,\ \xi_2(1) = 4.5 \end{aligned}$$

where r_t is the average prime rate, br_t the Aaa corporate bond rate, x_t is the industrial production index, i_t the three-month bill rate, d_t total bank deposits and q_t commercial loans. q_t and r_t are the endogenous variables, br_t, x_t, i_t and d_t are taken as at least weakly exogenous and no equation for these variables is explicitly estimated. Given that there are two omitted instruments in each equation one over-identifying restrictions is imposed both in the demand and in the supply equation. The validity of such restrictions is tested via the Anderson-Rubin tests($\xi_1(1)$ *and* $\xi_2(1)$), that leads to rejection of the restrictions at the

5 per cent level in both cases, but does not lead to rejection of the restrictions for the first equation ad the one per cent level. On the basis of this evidence, some (weak) support of the data on the chosen specification could be claimed.

The testing procedure is based by estimating as a statistical model a reduced form that projects all endogenous variables on all exogenous variables:

$$\begin{bmatrix} q_t \\ r_t \end{bmatrix} = \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} & \pi_{14} & \pi_{15} \\ \pi_{21} & \pi_{22} & \pi_{23} & \pi_{24} & \pi_{25} \end{bmatrix} \begin{bmatrix} 1 \\ br_t \\ x_t \\ d_t \\ i_t \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix}$$

and then by testing if the ten parameters in the unrestricted reduced form can be validly restricted to the eight in the structural model

$$\begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} & \pi_{14} & \pi_{15} \\ \pi_{21} & \pi_{22} & \pi_{23} & \pi_{24} & \pi_{25} \end{bmatrix} = \begin{bmatrix} 1 & \gamma_{12} \\ 1 & \gamma_{22} \end{bmatrix}^{-1} \begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{13} & 0 & 0 \\ \delta_{21} & 0 & 0 & \delta_{24} & \delta_{25} \end{bmatrix}$$

Spanos notes that estimation of the statistical model (i.e. the implicit unrestricted reduced form) yields:

$$\begin{aligned} q_t &= \underset{(21.05)}{-128.20} - \underset{(0.810)}{3.007} i_t + \underset{(1.236)}{7.078} br_t + \underset{(0.156)}{0.497} x_t + \underset{(0.011)}{0.281} d_t + \hat{u}_{1t} \\ r_t &= \underset{(3.02)}{1.864} + \underset{(0.116)}{0.771} i_t + \underset{(0.178)}{0.763} br_t + \underset{(0.022)}{0.008} x_t - \underset{(0.001)}{0.005} d_t + \hat{u}_{2t} \end{aligned}$$

where 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 is then considered allowing for a richer dynamic structure (two lags) in the reduced form, such dynamic specification is shown to provide a much better statistical model for the data than the static reduced from.

$$
\begin{bmatrix} q_t \\ r_t \end{bmatrix} = \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} & \pi_{14} & \pi_{15} \\ \pi_{21} & \pi_{22} & \pi_{23} & \pi_{24} & \pi_{25} \end{bmatrix} \begin{bmatrix} 1 \\ br_t \\ x_t \\ d_t \\ i_t \end{bmatrix} +
$$

$$
+\sum_{i=1}^{2} \begin{bmatrix} a_{11,i} & a_{12,i} \\ a_{21,i} & a_{22,i} \end{bmatrix} \begin{bmatrix} q_{t-i} \\ r_{t-i} \end{bmatrix} +
$$

$$
\sum_{i=1}^{2} \begin{bmatrix} b_{11,i} & b_{12,i} & b_{13,i} & b_{14,i} \\ b_{21,i} & b_{22,i} & b_{23,i} & b_{24,i} \end{bmatrix} \begin{bmatrix} br_{t-i} \\ x_{t-i} \\ d_{t-i} \\ i_{t-i} \end{bmatrix} +
$$

$$
\begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix}
$$

Of course, the adopted structural model implies many more over-identifying restrictions than the initial one. In fact, we have

$$
\begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} & \pi_{14} & \pi_{15} \\ \pi_{21} & \pi_{22} & \pi_{23} & \pi_{24} & \pi_{25} \end{bmatrix} = \begin{bmatrix} 1 & \gamma_{12} \\ 1 & \gamma_{22} \end{bmatrix}^{-1} \begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{13} & 0 & 0 \\ \delta_{21} & 0 & 0 & \delta_{24} & \delta_{25} \end{bmatrix}
$$

$$
\begin{bmatrix} a_{11,i} & a_{12,i} \\ a_{21,i} & a_{22,i} \end{bmatrix} = 0
$$

$$
\begin{bmatrix} b_{11,i} & b_{12,i} & b_{13,i} & b_{14,i} \\ b_{21,i} & b_{22,i} & b_{23,i} & b_{24,i} \end{bmatrix} = 0
$$

When tested, the validity of these thirteen restrictions both on the demand and supply equations is overwhelmingly rejected. 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.

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

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

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 (1)can be solved using standard numerical techniques (see, for example, Sims, 2002), and the solution can be expressed as:

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

where the matrices $\mathbf{A}_0, \mathbf{A}_1$,and $\mathbf{R}$ contain convolutions of the underlying model structural parameters. Note that, the solution is naturally represented as a VAR, of course it is a VAR potentially with stochastic singularity, as the dimension of the vector of shocks is typically smaller than that of the vector of variables included in the VAR. However, this problem is promptly solved by adding the appropriate number of measurement errors.

Recent Model Evaluation of DSGE models exploits the fact that a solved RBC model is a statistical model, in particular a VAR.

In general, the solved RBC model could be represented as a (structural) VAR[2]:

$$\begin{aligned}
\mathbf{Z}_t &= \mathbf{\Phi}_0^* (\theta) + \mathbf{\Phi}_1^* (\theta) \mathbf{Z}_{t-1} + ... + \mathbf{\Phi}_p^* (\theta) \mathbf{Z}_{t-p} + \mathbf{u}_t^* && (2) \\
\mathbf{u}_t^* &\sim N(\mathbf{0}, \Sigma_u^* (\theta)) \\
\mathbf{Z} &= \mathbf{X}\mathbf{\Phi}^* (\theta) + \mathbf{u}^* \\
\underset{Txn}{\mathbf{u}^*} &= \begin{bmatrix} \underset{Tx1}{\mathbf{u}_1^*} & ... & \underset{Tx1}{\mathbf{u}_n^*} \end{bmatrix} && (3) \\
\underset{Txn}{\mathbf{Z}} &= \begin{bmatrix} \underset{Tx1}{\mathbf{Z}_1} & ... & \underset{Tx1}{\mathbf{Z}_n} \end{bmatrix} && (4) \\
\underset{Tx(np+1)}{\mathbf{X}} &= \begin{bmatrix} \mathbf{X}_1' \\ \\ \mathbf{X}_T' \end{bmatrix}, \\
\underset{1x(np+1)}{\mathbf{X}_t'} &= \begin{bmatrix} 1, \underset{1xn}{\mathbf{Z}_{t-1}'} ... \underset{1xn}{\mathbf{Z}_{t-p}'} \end{bmatrix} && (5) \\
\underset{(np+1)xn}{\mathbf{\Phi}^* (\theta)} &= \begin{bmatrix} \underset{nx1}{\mathbf{\Phi}_0^* (\theta)}, \underset{nxn}{\mathbf{\Phi}_1^* (\theta)}, ..., \underset{nxn}{\mathbf{\Phi}_p^* (\theta)} \end{bmatrix}',
\end{aligned}$$

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

[2] In fact, solved DSGE model often generates a restricted MA representation for the vector of n variables of interest, that can be approximated by a VAR of finite order p.

unrestricted VAR. Note that following this logic 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

$$\begin{aligned} \mathbf{Z}_t &= \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{Z}_{t-1} + ... + \mathbf{\Phi}_p \mathbf{Z}_{t-p} + \mathbf{u}_t \qquad (6) \\ \mathbf{u}_t &\sim N(\mathbf{0}, \Sigma_u) \\ \mathbf{Z} &= \mathbf{X\Phi} + \mathbf{u} \\ \underset{(np+1)xn}{\mathbf{\Phi}} &= \left[\underset{nx1}{\mathbf{\Phi}_0}, \underset{nxn}{\mathbf{\Phi}_1}, ..., \underset{nxn}{\mathbf{\Phi}_p} \right]', \qquad (7) \end{aligned}$$

where:

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

the DSGE restrictions are imposed on the VAR by defining:

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

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:

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

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:

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

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:

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

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 takes the Lucas and Sims critique very seriously but 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 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 modification in the structure of the statistical model 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?

There are a number of potential sources of mis-specification for the model derived VAR. An obvious candidate are all those variables that are related to the mis-specification of the theoretical model, but there are also all those variables that are not theory related but are important to model the actual behaviour of policy makers. Think for example of the commodity price index and the modelling of the behaviour of monetary policy authority. It is by now common wisdom that the inclusion of this variable in a VAR to identify monetary policy shocks has been deemed important to model correctly the information set of the monetary policy maker when forecasting inflation and to fix the "price- puzzle"[3] in VAR based analysis of the monetary transmission mechanism. 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), this 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.

There are alternatives to the use of a VAR as a benchmark. We propose to address the limited information problem by combining traditional VAR analysis with recent developments in factor analysis for large data sets. We shall use a factor-augmented VAR (FAVAR) as the relevant statistical model to conduct model evaluation. 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. This approach has been successfully employed

[3] When impulse responses analysis is conducted in simple VAR models containing macro and monetary variables the response of prices to an innovation in interest rates gives rise to the 'price puzzle': prices increase significantly after an interest rate hike. The 'price puzzle' has been attributed to mis-specification of the small VARs. Suppose that there exists a leading indicator for inflation to which the Fed reacts. If such a leading indicator is omitted from the VAR, then we have an omitted variable positively correlated with inflation and interest rates. Such omission 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 1996) that the inclusion of a Commodity Price Index in the VAR solves the 'price puzzle'.

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

to forecast macroeconomic time series and in particular inflation. As a natural extension of the forecasting literature, Bernanke and Boivin (2003), Bernanke, Boivin and Eliasz(2005) proposed to exploit these factors in the estimation of VAR. A FAVAR benchmark for the evaluation of a DSGE model will take the following specification:

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

where $\mathbf{Z}_t$ are the 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 $\mathbf{Z}_t$. 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].

Interestingly what has instead already happened is that FAVAR have been interpreted as the reduced form of a DSGE model. This result has been achieved by removing the assumption that economic variables included in a DSGE are properly measured by a single indicator and by treating theoretical concepts of the model as partially observed to use the information set in factors to map them (Boivin and Giannoni,2005). This approach makes a FAVAR the reduced form a DSGE model, although the restrictions implied by DSGE model on a general FAVAR are very difficult to trace and model evaluation becomes even more difficult to implement. In fact, a very tightly parameterized theory model can have a very highly parameterized reduced form if one is prepared to accept that the relevant theoretical concept 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.

3 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.

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

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

$$\begin{aligned}
\tilde{x}_t &= E_t\tilde{x}_{t+1} - \frac{1}{\tau}(\tilde{R}_t - E_t\tilde{\pi}_{t+1}) + (1-\rho_G)\tilde{g}_t + \rho_z\frac{1}{\tau}\tilde{z}_t & (8)\\
\tilde{\pi}_t &= \beta E_t\tilde{\pi}_{t+1} + \kappa\,(\tilde{x}_t - \tilde{g}_t) & (9)\\
\tilde{R}_t &= \rho_R\tilde{R}_{t-1} + (1-\rho_R)(\psi_1\tilde{\pi}_t + \psi_2\tilde{x}_t) + \epsilon_{R,t} & (10)\\
\tilde{g}_t &= \rho_g\tilde{g}_{t-1} + \epsilon_{g,t} & (11)\\
\tilde{z}_t &= \rho_z\tilde{z}_{t-1} + \epsilon_{z,t} & (12)
\end{aligned}$$

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.

3.1 Solving the DSGE Model

We solve the model by applying the solution algorithm proposed by Sims (2002). We define the vector of variables as $\tilde{Z}_t = \begin{pmatrix} \tilde{x}_t & \tilde{\pi}_t & \tilde{R}_t & \tilde{R}^*_t & \tilde{g}_t & \tilde{z}_t & E_t\tilde{x}_{t+1} & E_t\tilde{\pi}_{t+1} \end{pmatrix}$ and the vector of shocks as $\epsilon_t = \begin{pmatrix} \epsilon_{R,t} & \epsilon_{g,t} & \epsilon_{z,t} \end{pmatrix}$. We can therefore recast the previous set of equations, (8) - (12), 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

$$\Gamma_0\tilde{Z}_t = C + \Gamma_1\tilde{Z}_{t-1} + \Psi\epsilon_t + \Pi\eta_t \tag{13}$$

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

As a solution to (13), we obtain the following policy function

$$\tilde{Z}_t = T(\theta)\tilde{Z}_{t-1} + R(\theta)\epsilon_t \tag{14}$$

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

$$\Delta \ln x_t = \ln\gamma + \Delta\tilde{x}_t + \tilde{z}_t \tag{15}$$

$$\Delta \ln P_t = \ln\pi^* + \tilde{\pi}_t \tag{16}$$

$$\ln R_t = 4[(\ln R^* + \ln\pi^*) + \tilde{R}_t] \tag{17}$$

which can be also cast into matrices as

$$Y_t = \Lambda_0(\theta) + \Lambda_1(\theta)\tilde{Z}_t + v_t \tag{18}$$

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 completeness, 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 (14) and (18) 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.

In what follows, we will sketch the main steps of the mixed estimation which combines the data information with the prior information deriving from the DSGE model. The measurement, (18), and the transition, (14), equation can be used to derive a sample of artificial data which are theory driven. We can therefore think of them as a set of dummy observations which can be added to the observables data as in Sims and Zha (1998)[6] to derive a prior distribution for the VAR coefficients. Furthermore, such a prior would be conjugate and that is re;event 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

$$P(\Phi, \Sigma_e, \theta \mid Y) = P_\Phi(\Phi, \Sigma_e \mid \theta, Y) \times P_\theta(\theta \mid Y), \tag{19}$$

where the first component can be easily calculated by using the conjugacy property of the DSGE-based prior while the second one, $P(\theta \mid Y)$, will be derived by

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

recalling MCMC methods. In particular, following DS, the Metropolis-Hastings will be employed to approximate the posterior.

3.2 Likelihood Function

The statistical benchmark proposed by DS to evaluate the DSGE model is an unrestricted VAR process. The variables included in the unrestricted VAR analysis are as above, $Y_t = (\Delta \ln x_t, \Delta \ln P_t, \ln R_t)'$. We consider a VAR with p lags where the dimension of Y_t is $(m, 1)$ while the dimension of the stacked vector X_t is $(k, 1)$ where $k = mp + 1$. The vector of innovations E_t, conditional on X_t, follow a multivariate Gaussian distribution $N_m (0, \Sigma_e)$.

$$\begin{aligned} Y_t &= \Phi_0 + \Phi_1 Y_{t-1} + \ldots \Phi_p Y_{t-p} + E_t, &(20)\\ &= \Phi X_t + E_t, &(21) \end{aligned}$$

where $\Phi = [\Phi_0 \mid \Phi_1 \mid \ldots \Phi_p]$ is of dimension (m, k) and the likelihood function, conditional on X_0, reads

$$\mathcal{L}(Y; \Phi, \Sigma_e) = (2\pi)^{-mT/2} |\Sigma_e|^{-T/2} \exp\left(-\frac{1}{2} tr\left(\left(\left(\Phi - \hat{\Phi}\right)(X'X)\left(\Phi - \hat{\Phi}\right)' + \hat{S}\right)\Sigma_e^{-1}\right)\right), \quad (22)$$

where we can write the likelihood function by recalling sufficient statistics such as $\hat{\Phi}' = (X'X)^{-1}(X'Y)$ and $\hat{S} = Y'Y - (Y'X)(X'X)^{-1}(X'Y)$.

Such a model represents an approximation to the MA representation of the theoretical model which is reliable as long as the number of lags in the unrestricted VAR is sufficiently large. However VAR models quickly become overparameterized as the number of lags increases which also deteriorates their forecasting performance.[7]

3.3 Prior Distribution

Following the approach by Sims (1996) about the use of dummy observations to impose a prior distribution on the set of coefficients, DS have assumed that such dummy observations could be derived from artificial data based on the simulation of the theoretical model such as the DSGE model highlighted above. We first write the likelihood function for a set of artificial data which are supposed to follow the same process as (21). The functional form of the likelihood is equivalent to (22), which is modified by the use of a Jeffreys prior: as we show in the appendix, this would lead to a proper DSGE-based prior to be used in

[7] This is why, as it was set forth by Zellner, a set of restrictions could turn out to be useful even if they are wrong. In particular, Bayesian VARs, by incorporating the shrinkage principle, have been shown to be useful in forecasting.

the posterior analysis.

$$\pi_1\left(\Phi,\Sigma_e \mid \tilde{Y},\tilde{X}\right) \propto (2\pi)^{-m\tilde{T}/2}\left|\Sigma_e\right|^{-(\tilde{v}+m+1)/2}\exp\left(-\frac{1}{2}tr\left(\tilde{S}\Sigma_e^{-1}\right)\right)\ldots \times\left|\Sigma_e\right|^{-k/2}\exp\left(-\frac{1}{2}tr\left(\left(\Phi-\tilde{\Phi}\right)\left(\tilde{X}'\tilde{X}\right)\left(\Phi-\tilde{\Phi}\right)'\Sigma_e^{-1}\right)\right) \tag{23}$$

where $\tilde{\Phi}' = \left(\tilde{X}'\tilde{X}\right)^{-1}\left(\tilde{X}'\tilde{Y}\right)$ and $\tilde{S} = \tilde{Y}'\tilde{Y} - \left(\tilde{Y}'\tilde{X}\right)\left(\tilde{X}'\tilde{X}\right)^{-1}\left(\tilde{X}'\tilde{Y}\right)$; here $\tilde{T} = \lambda T$ is the number of the artificial data set and $\tilde{v} = \tilde{T} - k$ are the relevant degrees of freedom in the likelihood function. Such a distribution is then the kernel of a Normal/Inverted-Wishart pdf which is a conjugate prior and therefore, conditional on the artificial data and in particular on θ, we will end up with a very tractable posterior distribution for the VAR coefficients, (Φ, Σ_e). The constant of integration and the proper prior distribution are fully derived in the appendix.

3.4 Posterior Distribution

In defining the posterior distribution we have to consider both the VAR coefficients and the DSGE model parameters. The joint posterior distribution with respect to (Φ, Σ_e, θ) conditional on having observed Y can be written as in (19). The first argument, $P(\Phi, \Sigma_e \mid \theta, Y)$, is what we find by combining the likelihood function with the prior based on the artificial data which are conditional on the set of parameters θ while the other one, $P(\theta \mid Y)$, is the posterior distribution of θ.

By combining the likelihood function (22) and the proper prior which derives from (23) we get the kernel of the posterior

$$P_\Phi(\Phi,\Sigma_e \mid \theta, Y) \propto \mathcal{L}(Y;\Phi,\Sigma_e) \times \pi_1\left(\Phi,\Sigma_e \mid \tilde{Y},\tilde{X}\right) \tag{24}$$

where we use the fact that artificial data are generated conditional to θ.[8]

Since we are in a conjugate prior case, we have a simple representation of the posterior distribution which can thus be partitioned in the product of a Inverted-Wishart distribution for Σ_e and Normal distribution for $\Phi \mid \Sigma_e$:

$$\Sigma_e \sim IW\left(\bar{S},\bar{v}\right), \tag{25}$$

$$\Phi \mid \Sigma_e \sim N_{m,k}\left(\bar{\Phi},\Sigma_e \otimes \bar{H}^{-1}\right), \tag{26}$$

where the elements entering in the two distributions are a function of the moments of the data and those of the DSGE model. In particular, to better gauge

[8] In the application we use population moments instead of drawing artificial data in order to eliminate stochastic uncertainty. However, these population moments critically depends on the θ.

how data and model moments mix each other we can look at the posterior mean of the VAR coefficients, Φ :

$$\bar{\Phi} = \bar{H}^{-1}\left(\lambda T\tilde{H}\tilde{\Phi} + T\hat{H}\hat{\Phi}\right) \tag{27}$$

which, given $\tilde{T} = \lambda T$, is governed by the parameter λ. Such a parameters can be thought of 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 since the prior would be flat.[9] Alternatively, as $\lambda \to \infty$ we have the posterior driven by the DSGE model only.

We have discussed the posterior of the VAR coefficients. We now want to explore the posterior of the DSGE model parameters, $P(\theta \mid Y)$, which can be rewritten as

$$P_\theta(\theta \mid Y) \propto P_Y(Y \mid \theta)\, \pi_2(\theta). \tag{28}$$

We therefore need a set of prior distributions for each structural parameter, $P(\theta)$, and the likelihood of the data given θ. The latter density can be obtained by integrating out (Φ, Σ_e) from the posterior (24). Here, we first describe the likelihood function, $P(Y \mid \theta)$, and then we sketch the steps of the random walk MH algorithm we will implement in the computation. By following DS, the marginal data density conditional on θ reads

$$P_Y(Y \mid \theta) = \int P(\Phi, \Sigma_e \mid \theta, Y)\, d\Phi d\Sigma_e, \tag{29}$$

and it has an analytical formula since we have a conjugate Normal/Inverted-Wishart prior. The prior distribution for θ is assumed to be a product of independent priors and it is summarized in the appendix. The shape of each prior distribution follows DS and they are consistent with the current literature on the estimation of DSGE model.

To simulate from the posterior distribution of θ we proceed as follows[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;

[9] The Jeffrey's prior we used for the DSGE based prior.
[10] Details of the derivation of the relevant posterior are described in Appendix C.

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.

4 DSGE Model Evaluation with a Statistically Identified Model

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.

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 (**??**).

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:

$$X_t = \Lambda^f \mathbf{F}_t + e_t \tag{30}$$

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.

5 Empirical Results

5.1 The Data

We analyse the DSGE-VAR model proposed by Del Negro and Schorfheide (2004) based on U.S. quarterly data from 1955:III to 2001:III.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]

On the basis of the 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.

5.2 The DSGE-VAR

We consider a benchmark DSGE-VAR model that replicates the results reported in Del Negro and Schorfheide (2004). We estimate a DSGE-VAR over the sample 1981-2001, considering the DSGE model described in section 2 and a four-order

[11] We extract factors by using principal components. We limit the number of factor 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.

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 grid $\Lambda = \{0.20, 0.60, 1, 1.4, 1.8, 10, Inf\}$. 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

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....".

6 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.

[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.

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
	χ^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
	χ^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 autocorrelation 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.

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

[13] See Godfrey(1988).

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:

$$\begin{aligned} RMSE^{y} &= \sqrt{\frac{1}{16}\sum_{h=1}^{16}\left(y_{t+h}-\hat{y}_{t+h|t}\right)^{2}} \qquad (31) \\ y &= \Delta \ln x_t, \Delta \ln P_t, \ln R_t, \\ t &= 1997:4 \end{aligned}$$

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.

7 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:

$$Y_t = B_0 X_t + B_1 F_t + E_t \tag{32}$$

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 (32) can be re-written in a more compact form as follows:

$$Y_t = BW_t + E_t \tag{33}$$

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.

7.1 Prior distribution

The full prior on the coefficients in (33) 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, $\left(\tilde{Y}_t, \tilde{X}_t\right)$,[14] 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

$$\begin{bmatrix} B_0 \\ B_1 \end{bmatrix} \mid \Sigma_e \sim N\left(\begin{bmatrix} \tilde{B}_0 \\ 0 \end{bmatrix}, \Sigma_e \otimes \begin{bmatrix} \left(\tilde{X}_t \tilde{X}_t'\right)^{-1} & 0 \\ 0 & \left(\tilde{F}_t \tilde{F}_t'\right)^{-1} \end{bmatrix}\right) \tag{34}$$

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.

7.2 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

$$P\left(\Phi, \Sigma_e, \theta \mid Y, F\right) = P_\Phi\left(\Phi, \Sigma_e \mid \theta, Y, F\right) \times P_\theta\left(\theta \mid Y, F\right), \tag{35}$$

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

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

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. 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.

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.

References

[1] An, Sungbae, and Frank Schorfheide (2006): "Bayesian Analysis of DSGE Models", *Working Paper, 06-5, Federal Reserve Bank of Atlanta.*

[2] Anderson, T.W. and H. Rubin (1949) "Estimation of the parameters of a single equation in a complete system of stochastic equations" Annals of Mathematical Statistics 20, 46—63.

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

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

[5] 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.

[6] Boivin, Jean, and Marc P. Giannoni (2005):"DSGE Models in a Data-Rich Environment," *Working Paper.*

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

[8] Christiano, Lawrence J. (2007): "Comment on M.Del Negro, F.Schorfheide, F.Smets and R.Wouters "On the Fit of New-Keynesian Models"", mimeo.

[9] 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.

[10] 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.

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

[12] 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*

[13] 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.*

[14] Forni, M. and L. Reichlin (1996), "Dynamic common factors in large cross-sections", *Empirical Economics*, 21, 27-42.

[15] Forni, M. and L. Reichlin (1998), "Let's get real: A dynamic factor analytical approach to disaggregated business cycle", *Review of Economic Studies*, 65, 453-474.

[16] Forni, M., Hallin, M., Lippi, M. and L. Reichlin (1999), "Reference cycles: The NBER methodology revisited", mimeo.

[17] Forni, M., Hallin, M., Lippi, M. and L. Reichlin (2000), "The generalized factor model: identification and estimation", *The Review of Economic and Statistics*,

[18] 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

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

[20] Godfrey L.G.(1988) "Misspecification Tests in Econometrics", Cambridge University Press

[21] Hendry, D. F. (1995). *Dynamic Econometrics.* Oxford: Oxford University Press.

[22] Jarque C.M. and Bera A.K.(1980) "Efficient tests for normality, homoscedasticity and serial independence of regression residuals", *Economic Letters*, 6, 255-259

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

[24] Maddala (1988) *Introduction to Econometrics*, Mac Millan, New York, NY

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

[26] Sims, C. A. (1980). 'Macroeconomics and Reality'. *Econometrica*, 48: 1-48.

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

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

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

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

[31] 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.

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

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

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

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

[36] White H.(1980) "A heteroscedastic-consistent covariance matrix estimator and a direct test for heteroscedasticity", Econometrica, 48, 817-838.

[37] Woodford M.(2003) "Interest and Prices", Princeton University Press

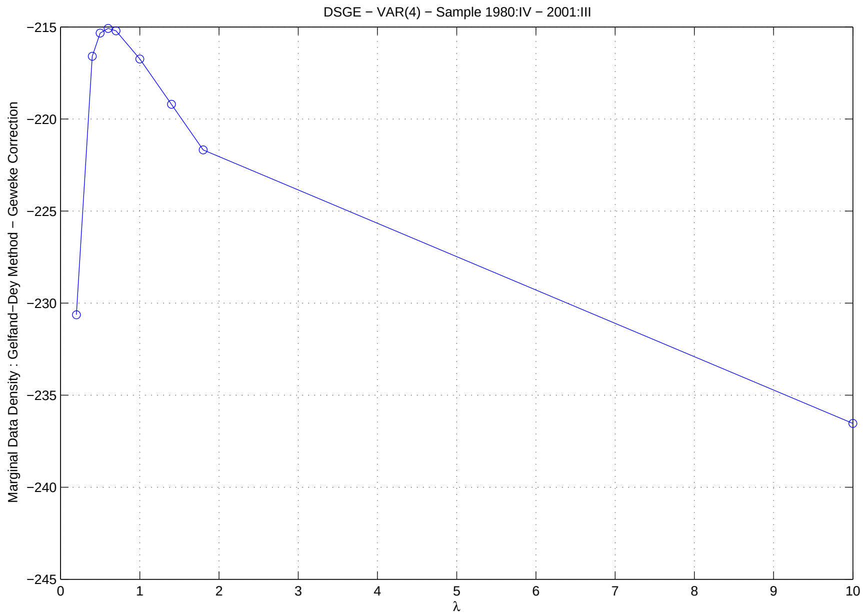

Figure 1: The optimal λ in the DSGE-VAR

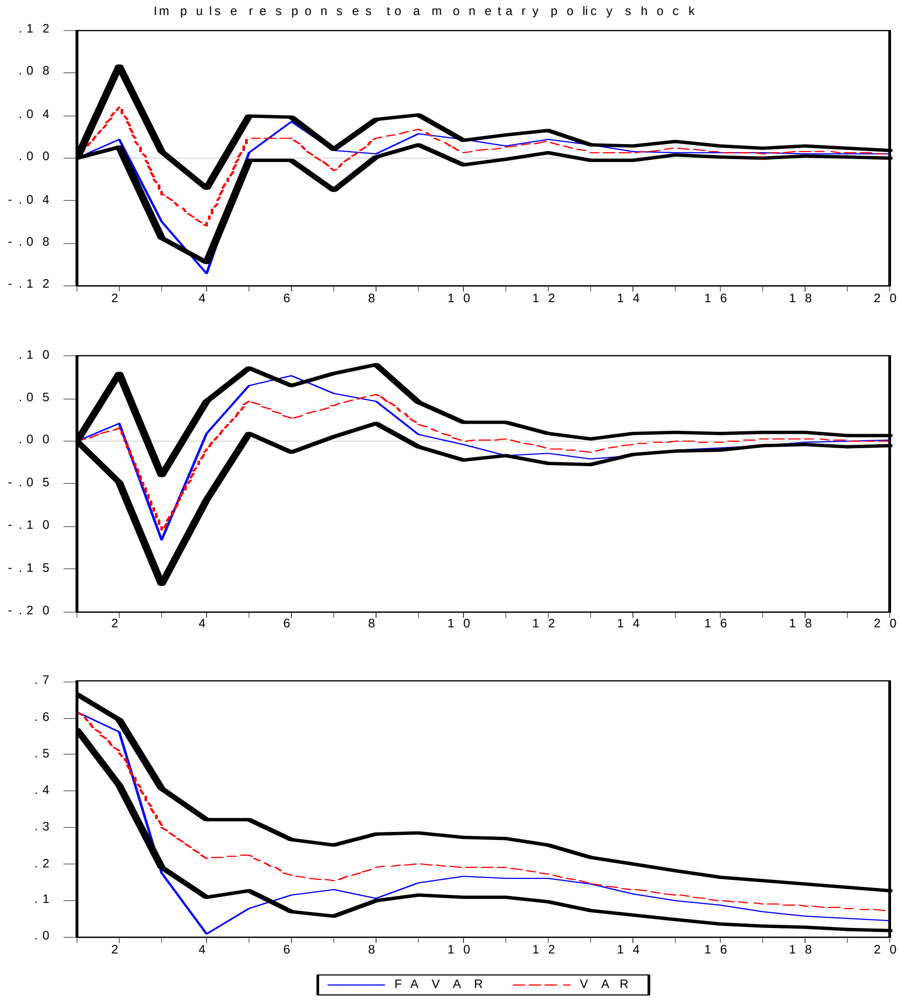

Figure 2:Responses of quarterly inflation, quarterly GDP growth and monetary policy rates to a monetary policy shock in the VAR and in the FAVAR(4)

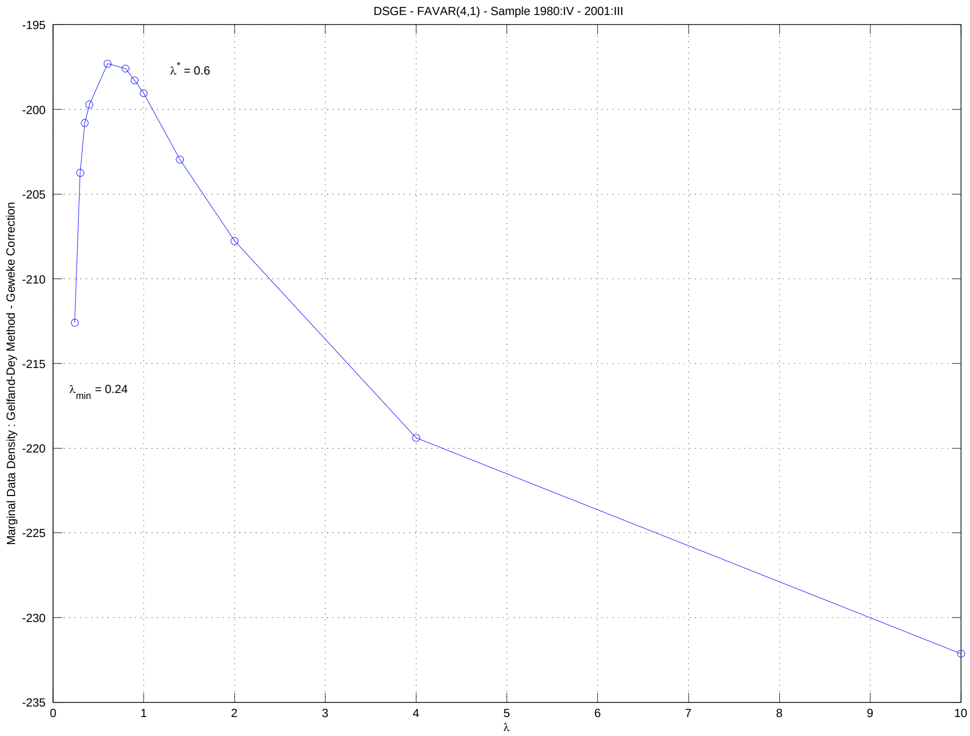

Figure 3: the optimal λ in a DSGE-FAVAR

9 Appendix A : The Sims Representation of our simple model

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

$$\tilde{x}_t = E_t\tilde{x}_{t+1} - \frac{1}{\tau}(\tilde{R}_t - E_t\tilde{\pi}_{t+1}) + (1-\rho_G)\tilde{g}_t + \rho_z\frac{1}{\tau}\tilde{z}_t \quad (36)$$

$$\tilde{\pi}_t = \beta E_t\tilde{\pi}_{t+1} + \kappa\,(\tilde{x}_t - \tilde{g}_t) \quad (37)$$

$$\tilde{R}_t = \rho_R\tilde{R}_{t-1} + (1-\rho_R)(\psi_1\tilde{\pi}_t + \psi_2\tilde{x}_t) + \epsilon_{R,t} \quad (38)$$

$$\tilde{g}_t = \rho_g\tilde{g}_{t-1} + \epsilon_{g,t} \quad (39)$$

$$\tilde{z}_t = \rho_z\tilde{z}_{t-1} + \epsilon_{z,t} \quad (40)$$

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

$$\Gamma_0\widetilde{\mathbf{Z}}_t = \Gamma_1\widetilde{\mathbf{Z}}_{t-1} + C + \Psi\epsilon_t + \Pi\eta_t \quad (41)$$

The results is achieved as follows:

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

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

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

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

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 transfer	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	iIP: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 mat	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 in	2	1
LHELX	EMPLOYMENT: RATIO; HELP-WANTED ADS:NO. UNEMPLOYED CLF	Help wanted/en	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

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)

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

$$Y_t = \Phi X_t + E_t, \tag{42}$$

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

$$\begin{aligned} \tilde{Y}_t &= \Lambda_0(\theta) + \Lambda_1(\theta)\tilde{Z}_t + V_t, & (43)\\ \tilde{Z}_t &= T(\theta)\tilde{Z}_{t-1} + R(\theta)U_t, & (44) \end{aligned}$$

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 (43) and (44) 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 (42)

$$\tilde{Y}_t = \Phi \tilde{X}_t + E_t, \tag{45}$$

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 (45).

11.2 Compute DSGE Moments

Given the state-space representation in (43) and (44), the unconditional variance for $\tilde{Y}_t$ and $\tilde{Z}_t$ are

$$\begin{aligned} \Sigma_{z,z} &= T\Sigma_{z,z}T' + R\Sigma_{u,u}R' && (46) \\ \Sigma_{y,y} &= \Lambda_0\Lambda_0' + \Lambda_1\Sigma_{z,z}\Lambda_1' + \Sigma_{v,v} + \Lambda_1 R\Sigma_{u,v} + \Sigma_{u,v}' R'\Lambda_1' && (47) \end{aligned}$$

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

$$\begin{aligned} \Sigma_{z,z}(k) &= T^k \Sigma_{z,z}(k-1) && (48) \\ \Sigma_{y,y}(k) &= \Lambda_0\Lambda_0' + \Lambda_1\Sigma_{z,z}(k)\Lambda_1' + \Lambda_1\left(T^k\right)R\Sigma_{u,v}. && (49) \end{aligned}$$

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

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

The likelihood function for the artificial data in (**??**) reads

$$\mathcal{L}\left(\tilde{Y};\Phi,\Sigma_e\right) = (2\pi)^{-mT/2}\left|\Sigma_e\right|^{-T/2}\exp\left(-\frac{1}{2}tr\left(\left(\left(\Phi-\tilde{\Phi}\right)\left(\tilde{X}'\tilde{X}\right)\left(\Phi-\tilde{\Phi}\right)+\tilde{S}\right)\Sigma_e^{-1}\right)\right), \tag{50}$$

where the sufficient statistics are,

$$\begin{aligned} \tilde{\Phi} &= \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{Y} && (51) \\ \tilde{S} &= \tilde{Y}'\tilde{Y} - \tilde{Y}'\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{Y} && (52) \end{aligned}$$

which can be also specified in terms of population moments

$$\begin{aligned} \tilde{\Phi} &= \Sigma_{x,x}^{-1}\Sigma_{x,y} && (53) \\ \tilde{S} &= \Sigma_{y,y} - \Sigma_{x,y}'\Sigma_{x,x}^{-1}\Sigma_{x,y} && (54) \end{aligned}$$

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

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

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

$$\pi_0 = |\Sigma_e|^{-\frac{m+1}{2}} . \tag{55}$$

By combining (50) and (55) we get the kernel of the distribution

$$\pi_1 \propto \mathcal{L}\left(\tilde{Y} \mid \Phi, \Sigma_e\right) \times \pi_0, \tag{56}$$

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

$$P_{\tilde{Y}}\left(\tilde{Y} \mid \theta\right) = (2\pi)^{-m\tilde{v}/2} \times \left|\tilde{S}\right|^{-\frac{\tilde{v}}{2}} \times \left|\tilde{H}\right|^{-\frac{m}{2}} \times \left[2^{m\tilde{v}/2} \times \pi^{m(m-1)/4} \times \Gamma_m\left(\tilde{v}\right)\right], \tag{57}$$

which is needed to have the DSGE-based prior distribution

$$\begin{aligned} \pi_1\left(\Phi, \Sigma_e \mid \tilde{Y}, \theta\right) &= \frac{\mathcal{L}\left(\tilde{Y} \mid \Phi, \Sigma_e, \theta\right) \times \pi_0}{P_{\tilde{Y}}\left(\tilde{Y} \mid \theta\right)} \\ &= \frac{(2\pi)^{-mT^*/2}}{(2\pi)^{-m\tilde{v}/2}} \times \frac{\left|\tilde{S}\right|^{\tilde{v}/2} \times \left|\tilde{H}\right|^{m/2} \times |\Sigma_e|^{-(T^*+m+1)/2}}{2^{m\tilde{v}/2} \times \pi^{m(m-1)/4} \times \Gamma_m\left(\tilde{v}\right)} \times \\ &\quad \exp\left[-\frac{1}{2} tr\left(\tilde{S}\Sigma_e^{-1}\right)\right] \times \exp\left[-\frac{1}{2} tr\left(\left(\Phi - \tilde{\Phi}\right)'\left(\Sigma_{x,x}\right)\left(\Phi - \tilde{\Phi}\right)\Sigma_e^{-1}\right)\right], \end{aligned} \tag{58}$$

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.

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

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

$$P_\Phi\left(\Phi, \Sigma_e \mid Y, \theta\right) \propto \mathcal{L}\left(Y \mid \Phi, \Sigma_e\right) \times \pi_1\left(\Phi, \Sigma_e \mid \tilde{Y}, \theta\right), \tag{59}$$

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

$$P_Y\left(Y \mid \theta\right)=\left(2\pi\right)^{-Tm/2}\times\frac{\left|\tilde{S}\right|^{\tilde{v}/2}}{\left|\bar{S}\right|^{\bar{v}/2}}\frac{\left|\tilde{H}\right|^{m/2}}{\left|\bar{H}\right|^{m/2}}\times\frac{\Gamma_m\left(\bar{v}\right)}{\Gamma_m\left(\tilde{v}\right)}\times 2^{m(\hat{v}+k)/2} \tag{60}$$

The proper posterior reads

$$\begin{aligned} P_\Phi\left(\Phi,\Sigma_e \mid Y,\theta\right) &= \left(2\pi\right)^{-mk/2}\times\left|\Sigma_e\right|^{-k/2}\times\exp\left[-\frac{1}{2}tr\left(\left(\Phi-\bar{\Phi}\right)'\bar{H}\left(\Phi-\bar{\Phi}\right)\Sigma_e^{-1}\right)\right]\dots \\ &\quad\times\frac{\left|\bar{S}\right|^{\bar{v}/2}\left|\bar{H}\right|^{m/2}\times\left|\Sigma_e\right|^{-(\hat{v}+T^*+m+1)/2}}{2^{m\bar{v}/2}\times\pi^{m(m-1)/4}\Gamma_m\left(\bar{v}\right)}\times\exp\left[-\frac{1}{2}tr\left(\bar{S}\Sigma_e^{-1}\right)\right] \end{aligned} \tag{61}$$

or equivalently

$$p\left(\Phi \mid \Sigma_e; Y, X\right) = N\left(\bar{\Phi},\Sigma_e\otimes\bar{H}^{-1}\right) \tag{62}$$

$$p\left(\Sigma_e \mid Y, X\right) = IW\left(\bar{S},\bar{v}\right) \tag{63}$$

where the posterior estimates are as follows

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

11.5 Metropolis-Hasting Algorithm

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

$$P\left(\Phi,\Sigma,\theta \mid Y\right)=P_\Phi\left(\Phi,\Sigma \mid Y,\theta\right)\times P_\theta\left(\theta \mid Y\right) \tag{64}$$

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

$$P_\theta\left(\theta \mid Y\right)\propto\mathbb{K}_\theta\left(\theta \mid Y\right)=P_Y\left(Y \mid \theta\right)\times\pi_2\left(\theta\right) \tag{65}$$

[16] where

$$\tilde{H} = \left(\tilde{X}'\tilde{X}\right)$$

$$\bar{v} = T+\tilde{T}-k$$

where $P_Y(Y \mid \theta)$ has been computed above and $\pi_2(\theta)$ is a set of independent prior distributions over each element of the vector of parameters θ; $\mathbb{K}_\theta(\theta \mid Y)$ is the kernel of the posterior. By combining the likelihood and the prior we don't have a closed form solution. We thus need to simulate draws out of the posterior distribution which is unknown. We follow Schorfheide (2000) and 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, $\tilde{\theta}$,

$$\tilde{\theta} = \theta_{s-1} + \left(\Sigma_H(\theta^*)\right)^{-1/2} N(0, I), \tag{66}$$

we construct an acceptance probability threshold

$$\alpha\left(\tilde{\theta}, \theta_{s-1}\right) = \min\left(1, \frac{\mathbb{K}_\theta\left(\tilde{\theta} \mid Y\right)}{\mathbb{K}_\theta(\theta_{s-1} \mid Y)}\right). \tag{67}$$

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 (66).

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.

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.