Priors from General Equilibrium Models for VARs: Forecasting and Identification*

Marco Del Negro
Research Department
Federal Reserve Bank of Atlanta[†]

Frank Schorfheide
Department of Economics
University of Pennsylvania[‡]

February 2002

Preliminary and Incomplete

JEL CLASSIFICATION: C11, C32, C53

KEY WORDS: Bayesian Analysis, DSGE Models, Forecasting, Vector Autoregressions

*Part of this research was conducted while the second author was visiting the Research Department of the Federal Reserve Bank of Atlanta, for whose hospitality he is grateful. The second author gratefully acknowledges financial support from the University Research Foundation of the University of Pennsylvania. The views expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

[†]1000 Peachtree Street NE, Atlanta GA 30309-4470. Email: Marco.DelNegro@atl.frb.org

[‡]3718 Locust Walk, Philadelphia, PA 19104-6297. Email: schorf@ssc.upenn.edu

Abstract

Ingram and Whiteman (1994) have shown that priors from DSGE models can be helpful in forecasting. However, Ingram and Whiteman's approach applies only geared toward trend-stationary, reduced-form VARs and a subset of DSGE models, namely those for which the log-linearized solution has a VAR representation. We generalize their approach so that it easily can be applied to any log-linearized DSGE model and to any VAR. We use a hierarchical prior that consists of a marginal distribution for DSGE model parameters θ and a distribution for the VAR parameters (Φ, Σ_u) conditional on θ. Markov-Chain-Monte-Carlo methods are used to compute a joint posterior distribution for θ and (Φ, Σ_u). We show that the posterior updating of θ is crucial for accurate forecasting. Finally, we use the DSGE model not only for forecasting, but also to achieve identification, thereby making the approach fully exploitable for policy analysis. We apply this framework to a bivariate VAR in inflation and output, using the prior information coming from Christiano and Eichenbaum (1992) liquidity effect model. We find that the VAR with DSGE priors forecasts better in several dimensions than an array of competitors, particularly as far as long run inflation is concerned.

1 Introduction

For vector autoregressions (VARs) to be useful in policy analysis they have to be successful in two dimensions: forecasting and identification. Dynamic stochastic general equilibrium (DSGE) models can help in both regards. There is a small literature (DeJong, Ingram, and Whiteman (1993), Ingram and Whiteman (1994), Schorfheide (1998)) that has used DSGE models to derive a prior distribution for the coefficients of a VAR. However, none of these papers makes an attempt to use the DSGE model also for identification purposes. Vice versa, there is a growing literature, e.g., Canova (2001) and Chang and Schorfheide (2000), that uses DSGE models to characterize a set of economically plausible impulse response patterns and then constructs a VAR identification scheme that is consistent with this pattern.[1] However, when estimating the reduced form coefficients of the VAR, these papers do not employ prior distributions that shrink the parameter estimates toward the restrictions embodied in DSGE models. Moreover, in both strands of the literature the DSGE models and their parameters are chosen *a priori*. The researcher does not explicitly learn from the data with parameterization is *a posteriori* the best for forecasting and identification.

This paper proposes a methodology that uses prior information from DSGE models to improve a VAR's forecasting performance and to achieve identification. In short, the approach works as follows. The researcher chooses one or more DSGE models and specifies a prior distribution for their parameters, which is mapped into a prior for the VAR parameters. Recent developments in Bayesian econometrics offer a set of techniques to combine prior information and data, and to obtain a joint posterior distribution for DSGE and VAR parameters. The techniques described in Canova and De Nicolo (2001), Faust (1998), Uhlig (1999), and a modification of Canova and De Nicolo's procedure proposed in this paper furnish a mapping from the joint posterior into VAR impulse response functions, thereby achieving identification. The single components of the approach are by no means new. The

[1] There are several additional procedures available to implement a DSGE model consistent identification of a VAR: Canova and De Nicolo (2001), Faust (1998), Uhlig (1999).

novelty of the paper – aside from some technical innovations that make it possible to generalize some existing results, and to ease the computational burden of the procedure – consists in pasting together these components into a viable strategy.

In terms of forecasting, it has been known for more than a decade that the performance of unrestricted VARs can be improved by the use of prior information in the estimation procedure. Bayesian vector autoregressions were first utilized for macroeconomic prediction in Doan, Litterman, and Sims (1984) and thereafter discussed, for instance, in Litterman (1984, 1986), Todd (1984), and Sims (1993). The popular Minnesota prior shrinks the VAR coefficients toward univariate unit root representations. While the Minnesota prior has been empirically successful, it lacks economic justification and ignores information about the interaction of the elements of the vector of endogenous variables. As an alternative, DeJong, Ingram, and Whiteman (1993) and Ingram and Whiteman (1994), proposed to derive prior distributions from DSGE models. For instance, the latter use a prototypical stochastic growth model to obtain a prior distribution and show that the VAR with their DSGE model prior outperforms the forecasts of an unrestricted VAR and attains approximately the same forecast error loss as a VAR with Minnesota prior.

Ingram and Whiteman (1994) estimate the VAR equation by equation rather than treating it as a multivariate system. Thus, they disregard potentially useful information of the DSGE model about correlations of parameters across equations. Ingram and Whiteman also assumed that the log-linear approximation of the DSGE model leads to a first-order VAR for the variables that are being forecasted. While this assumption was satisfied in their specific application, it will in many cases not be satisfied. In general, DSGE models have state-space representations that strictly speaking lead to infinite-order vector autoregressions. Moreover, Ingram and Whiteman's procedure does not *a posteriori* re-center the prior distribution to achieve the best fit of the VAR, nor is the DSGE model used for the purpose of identifying structural shocks in the estimated VAR. This paper extends and generalizes the Ingram and Whiteman approach to overcome the aforementioned shortcomings.

We will specify a hierarchical prior starting out with the prior distribution of the

DSGE model parameters and a tightness parameter that controls the relative weight of the VAR prior. Conditional on the DSGE model parameters we specify a prior for the VAR parameters without assuming that the log-linearized DSGE model is nested in the VAR. Our procedure is general enough to accommodate VARs with stationary and non-stationary endogenous variables. Markov-Chain Monte Carlo methods are used to generate draws from the joint posterior distribution of the VAR and DSGE model parameters.

Based on the joint posterior it is possible to simulate predictive densities for future values of the endogenous variables, and to compute and evaluate out-of-sample forecasts. To identify the VAR, we rotate the reduced-form VAR impulse responses for each draw from the joint posterior so that they qualitatively match the DSGE model's responses according to some criterion. The outcome of this procedure is a distribution of *identified* impulse-responses for the VAR, which can in turn be used for conditional forecasting exercises and policy analysis. A nice feature of this approach is that the bands for the identified impulse responses fully reflect the modeler's posterior uncertainty about the DSGE parameters and hence the set of economically plausible impulse response patterns. It is important to note that our procedure works just as well if one does not want to use the DSGE model for the sake of identification, and prefers to rely on linear restrictions on the variance covariance matrix of innovations (as in Bernanke (1986) and Sims (1986)).

The paper is organized as follows. The notation and setup of our hierarchical Bayes model is explained in Section **??**. Section 2 contains a brief description of the DSGE model that we are using to construct the prior distribution. Section 3 discusses DSGE model priors for VARs with trend-stationary endogenous variables whereas Section 3.3 generalizes the approach to VARs with non stationary endogenous variables. Empirical results for an output and price VAR are presented in Section 4 and Section 5 concludes.

2 A Simple Monetary DSGE Model

As an empirical application in this paper we will consider a bivariate VARs for output and the aggregate price level. We will derive prior distributions for the VARs from a cash-in-advance model in which households make their deposit decisions before observing the money supply shock and face portfolio management costs. This model was proposed by Christiano and Eichenbaum (1992) and its empirical adequacy was studied, for instance, in Nason and Cogley (1994) and Schorfheide (2000). To make this paper self-contained we briefly review the model specification.

The model economy consists of a representative household, a firm, and a financial intermediary. Output is produced according to a Cobb-Douglas production function

$$GDP_t = K_t^{\alpha}(A_t N_t)^{1-\alpha}, \tag{1}$$

where K_t denotes the capital stock (predetermined at the beginning of period t), N_t is the labor input and A_t is a labor augmenting technology.

The model economy is perturbed by two exogenous processes. Technology evolves according to

$$\ln A_t = \gamma + \ln A_{t-1} + \epsilon_{A,t}, \quad \epsilon_{A,t} \sim \mathcal{N}(0, \sigma_A^2). \tag{2}$$

The central bank lets the money stock M_t grow at rate $m_t = M_{t+1}/M_t$:

$$\ln m_t = (1-\rho)\ln m^* + \rho \ln m_{t-1} + \epsilon_{M,t}, \quad \epsilon_{M,t} \sim \mathcal{N}(0, \sigma_M^2). \tag{3}$$

Equation (3) can be interpreted as a simple monetary policy rule without feedbacks. Both (log) productivity and money stock are drifting random walk processes and introduce two stochastic trends into the model.

At the beginning of period t, the representative household inherits the entire money stock of the economy, M_t. We assume that in period $t-1$ the household has determined how much money D_t it will deposit at the bank in period t. These deposits earn interest at the rate $R_{H,t} - 1$. Let the household's cash holdings be denoted by $Q_t = M_t - D_t$ and let P_t be the aggregate price level.

After observing the period t shocks the household chooses consumption C_t, hours worked H_t, and next period's cash-holdings Q_{t+1} to maximize the sum of discounted expected future utility. It solves the problem

$$\begin{aligned}
\max_{\{C_t,H_t,M_{t+1},Q_{t+1}\}} \quad & I\!E_0\left[\sum_{t=1}^{\infty}\beta^t[(1-\phi)\ln C_t+\phi\ln(1-H_t-\tilde{p}_t)]\right] \qquad (4)\\
\text{s.t.} \quad & P_tC_t \le Q_t + W_tH_t\\
& Q_t \le M_t\\
& M_{t+1} = (Q_t + W_tH_t - P_tC_t) + R_{H,t}(M_t - Q_t) + F_t + B_t.
\end{aligned}$$

Adjustments to the level of cash-holdings reduce the time available for leisure. The household's portfolio management costs are given by

$$\tilde{p}_t = \alpha_1\left[\exp\left\{\alpha_2[\frac{Q_t}{Q_{t-1}} - m^*]\right\} + \exp\left\{-\alpha_2[\frac{Q_t}{Q_{t-1}} - m^*]\right\} - 2\right], \qquad (5)$$

The household receives its wage W_tH_t in cash which increases its money balance to $Q_t + W_tH_t$. Here W_t is the nominal hourly wage. The cash-in-advance constraint implies that all consumption purchases must be paid for with the accumulated cash balance. At the end of the period the household receives back its bank deposits inclusive of interest. Moreover, the net cash inflow of the firm and the financial intermediary are paid back as dividends (F_t and B_t) to the household.

The financial intermediary receives household deposits and a monetary injection X_t from the central bank, which it lends to the firm at rate $R_{F,t} - 1$. The bank solves the trivial problem

$$\begin{aligned}
\max_{\{B_t,L_t,D_t\}} \quad & I\!E_0\left[\sum_{t=1}^{\infty}\beta^{t+1}\frac{B_t}{C_{t+1}P_{t+1}}\right] \qquad (6)\\
\text{s.t.} \quad & B_t = D_t + R_{F,t}L_t - R_{H,t}D_t - L_t + X_t\\
& L_t \le X_t + D_t,
\end{aligned}$$

where $X_t = M_{t+1} - M_t$ is the monetary injection. In equilibrium $R_{F,t} = R_{H,t}$. Since the household cannot revise its deposit decision after a surprise change in the money growth rates, the additional cash has to be absorbed by the firm, which forces the nominal interest rate to fall and creates a liquidity effect. The portfolio-adjustments faced by the household increase the persistence of this liquidity effect. The drop in interest rates stimulates economic activity. In the long-run, however, output returns to its pre-intervention steady state level. Prices will be permanently higher.

After the realization of the time t shocks the firm starts production and hires labor services from the household. It uses the money borrowed from the financial intermediary to pay the wage bill. The firm chooses next period's capital stock K_{t+1}, labor demand N_t, loans L_t, and dividends F_t. Since households value a unit of nominal dividends in terms of the consumption it enables during period $t+1$, and firms and the financial intermediary are owned by households, date t nominal dividends are discounted by date $t+1$ marginal utility of consumption. Thus, the firm solves the problem

$$\begin{aligned} \max_{\{F_t, K_{t+1}, N_t, L_t\}} \quad & I\!\!E_0 \left[\sum_{t=0}^{\infty} \beta^{t+1} \frac{F_t}{C_{t+1} P_{t+1}} \right] \qquad (7) \\ \text{s.t.} \quad & F_t \leq L_t + P_t [K_t^\alpha (A_t N_t)^{1-\alpha} - K_{t+1} + (1-\delta) K_t] - W_t N_t - L_t R_{F,t} \\ & W_t N_t \leq L_t. \end{aligned}$$

The market clearing conditions for labor market, money market, and goods market are $H_t = N_t$, $P_t C_t = M_t + X_t$, and $C_t + (K_{t+1} - (1-\delta)K_t) = K_t^\alpha (A_t H_t)^{1-\alpha}$.

To solve the models, optimality conditions are derived for the maximization problems. The real variables are then detrended by the productivity A_t, the price level by M_t/A_t, and X_t, Q_t, and D_t are detrended by M_t. It can be shown that the system in the detrended variables has a deterministic steady state and can be log-linearized around it. A solution to this linear rational expectation system can be obtained by elimination of unstable roots according to the algorithm in Sims (1995). The log-linear approximation of the DSGE model only depends on $\kappa = \alpha_1 \alpha_2^2$. The structural parameters are stacked in the vector

$$\theta = [\alpha, \beta, \gamma, m^*, \rho, \phi, \delta, \kappa, \sigma_A, \sigma_M]'. \qquad (8)$$

Moreover, define $\epsilon_t = [\epsilon_{A,t}, \epsilon_{B,t}]'$.

The log-linearized DSGE model generates a joint probability distribution for aggregate output and prices, both in log-levels and in growth rates. The law of motion of the money stock can be rewritten as

$$\ln M_t = \ln m^* + \ln M_{t-1} + \sum_{j=0}^{\infty} \rho^j \epsilon_{M,t-j} \tag{9}$$

Define the vector of stochastic trends $z_{2,t} = [\ln A_t, \ln M_t]'$, which evolves according to

$$z_{2,t} = \mu_2 + z_{2,t-1} + C_2(L)\epsilon_t. \tag{10}$$

The drift μ_2 is equal to $[\gamma, \ln m^*]$, and $C_2(L)$ is the lag polynomial $\sum_{j=0}^{\infty} C_{2,j}(L)$. The $C_{2,j}$'s are 2×2 matrices such that Equation (10) is equivalent to Equations (2) and (9).

The law of motion for output and prices is of the form

$$\underbrace{\begin{bmatrix} \ln GDP_t \\ \ln P_t \end{bmatrix}}_{z_{1,t}} = \underbrace{\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} \ln A_t \\ \ln M_t \end{bmatrix}}_{z_{2,t}} + \underbrace{\begin{bmatrix} gdp^* \\ p^* \end{bmatrix}}_{\mu_1} + \underbrace{\begin{bmatrix} \widehat{gdp}_t \\ \widehat{p}_t \end{bmatrix}}_{C_1(L)\epsilon_t}. \tag{11}$$

Both output and prices are non-stationary due to the stochastic trends in technology and money. gdp^* and p^* are the log-steady state values of output and prices in the detrended DSGE model. $\widehat{gdp}_t$ and $\widehat{p}_t$ denote percentage deviations from steady state.[2] In the log-linear approximation of the equilibrium, these percentage deviations have an infinite-order moving average representation in terms of the structural shocks ϵ_t, denoted by $C_1(L)\epsilon_t$.

Upon differencing of $z_{1,t}$ we obtain a law of motion for output growth and inflation

$$\Delta z_{1,t} = \begin{bmatrix} \Delta \ln GDP_t \\ \Delta \ln P_t \end{bmatrix} = \Lambda[\mu_2 + C_2(L)\epsilon_t] + C_1(L)(1-L)\epsilon_t. \tag{12}$$

The differencing removes the stochastic trends such that output growth and inflation are stationary.

[2] In this model the steady state values could be absorbed into the initialization of the stochastic trends.

3 Constructing VAR Priors from the DSGE Model

Let y_t be the $n \times 1$ vector of endogenous variables. In the context of the application described in the previous section y_t either equals $z_{1,t}$ or $\Delta z_{1,t}$. The VAR model is of the form

$$y_t = c + \Phi_0 + \Phi_1 y_{t-1} + \ldots \Phi_p y_{t-p} + u_t, \quad u_t \sim \mathcal{N}(0, \Sigma_u), \tag{13}$$

where u_t is a vector of reduced-form disturbances. Let Y_T be the $T \times n$ matrix with rows y_t'. Let $k = 1 + np$, X_T be the $T \times k$ matrix with rows $x_t' = [1, y_{t-1}', \ldots, y_{t-p}']$, U_T be the $T \times n$ matrix with rows u_t', and $\Phi = [c, \Phi_0, \Phi_1, \ldots, \Phi_p]'$. The VAR can be expressed as

$$Y_T = X_T \Phi + U_T \tag{14}$$

with likelihood function

$$p(Y_T|\Phi, \Sigma_u) \propto |\Sigma_u|^{-T/2} \exp\left\{ -\frac{1}{2} tr[\Sigma_u^{-1}(Y_T - X_T\Phi)'(Y_T - X_T\Phi)] \right\} \tag{15}$$

conditional on observations $y_{1-p}, \ldots, y_0$.

In order to conduct Bayesian inference we will specify a hierarchical prior of the form

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

where θ is a vector of DSGE model parameters and λ is a hyperparameter that controls the weight of the prior relative to the sample information. The joint posterior distribution of Φ, Σ_u, θ, and λ can be written as

$$p(\Phi, \Sigma_u, \theta, \lambda|Y_T) = p(\Phi, \Sigma_u|\theta, \lambda)p(\theta, \lambda|Y_T). \tag{16}$$

The first factor represents the posterior of the reduced for VAR parameters Φ and Σ_u conditional on DSGE model parameters θ and hyperparameter λ. The second factor is the marginal posterior of θ and λ. Based on the posterior for Φ and Σ_u we can (numerically) obtain a predictive distribution for future y_{T+h}'s.

Let $\Sigma_u^{1/2}$ be the Cholesky decomposition of Σ_u. The reduced-form disturbances are related to a vector of structural shocks according to

$$u_t = \Sigma_u^{1/2}\Phi_* \epsilon_t = \Phi_\epsilon \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, I). \tag{17}$$

The transformation matrix Φ_* is orthonormal. An identification scheme for the structural shocks ϵ_t is a mapping from the VAR parameters Φ, Σ_u and the DSGE model parameters θ into the space of orthonormal $n \times n$ matrices

$$\Phi_* = \mathcal{I}(\Phi, \Sigma_u, \theta). \tag{18}$$

An identification scheme allows us to generate forecasts conditional on specific realizations of the structural shocks.

In order to construct a DSGE model prior $p(\Phi, \Sigma_u|\theta, \lambda)$ we will, roughly speaking, augment the actual data by $T^* = \lambda T$ artificial observations generated from the DSGE model. The hyperparameter λ reflects the ratio between the size of the simulated sample and the actual sample. This approach of using artificial or dummy observations to incorporate prior information is known as mixed estimation and originally due to Theil and Goldberger (1961). The use of dummy observations to induce priors in Bayesian VARs is quite common, e.g., Sims and Zha (1998). Rather than generating random observations $y_1^*, \ldots, y_{T_*}^*$ from Equation (11) or (12) and augmenting the actual data Y_T, that is, pre-multiplying the likelihood function by

$$p(\Phi, \Sigma_u|\theta, \lambda) \propto |\Sigma_u|^{-\lambda T/2} \exp\left\{ -\frac{1}{2} tr[\Sigma_u^{-1}(Y_{T_*}^* - X_{T_*}^* \Phi)'(Y_{T_*}^* - X_{T_*}^* \Phi)] \right\}, \tag{19}$$

we will take an asymptotic approach. Such an approach bears two advantages: (i) it save enormous computational burden, (ii) it prevents sampling error from the model to interfere with the construction of the prior distribution.[3]

3.1 VARs with Stationary Endogenous Variables

If the vector of endogenous variables y_t is composed of output growth and inflation, that is, $y_t = \Delta z_{1,t}$ according to the Notation of Section 3, then y_t is covariance

[3] If one does not take an asymptotic approach, one is left with a tradeoff between *(i)* and *(ii)*: sampling error can be virtually eliminated at the cost of producing a large number of draws at each step of the Metropolis-Hastings algorithm, making the overall procedure computationally very expensive.

stationary and the following moment matrices are well-defined:

$$\Gamma_{yy} = I\!\!E[y_t y_t'], \quad \Gamma_{yx} = I\!\!E[y_t x_t'], \quad \Gamma_{xx} = I\!\!E[x_t x_t'].$$

A law of large numbers for weakly dependent processes yields the conclusion that sample moments computed from the dummy observations converge to population moments as the sample size tends to infinity, for instance

$$\lim_{T_* \to \infty} \frac{1}{T} Y_{T_*}^{*\prime} Y_{T_*}^{*} = \Gamma_{yy}(\theta). \tag{20}$$

The limit matrix of cross moments is a function of the structural parameters.

We now replace the sample moments in (21) by scaled population moments and use the prior

$$p(\Phi, \Sigma_u | \theta, \lambda) \propto |\Sigma_u|^{-\frac{\lambda T + n + 1}{2}} \exp\left\{ -\frac{1}{2} tr[\lambda T \Sigma_u^{-1} (\Gamma_{yy} - \Phi' \Gamma_{yx}' - \Gamma_{xy} \Phi + \Phi' \Gamma_{xx} \Phi)] \right\}. \tag{21}$$

The hyperparameter λ determines the effective sample size for the artificial observations, which is λT: when λ is equal to (less than, greater than) 1 the prior will receive as much (less, more) weight in the posterior as the actual data.

In order to analyze the posterior distribution we use the following factorization

$$p(\Phi, \Sigma_u, \theta, \lambda | Y) = p(\Phi, \Sigma_u | Y, \theta, \lambda) p(\theta | Y, \lambda) p(\lambda | Y). \tag{22}$$

We assume that the parameter space of λ is finite $\Lambda = \{l_1, \ldots, l_q\}$. Thus, $p(\lambda|Y)$ reflects the posterior probabilities of λ. To reduce the computational burden we condition on the $\hat{\lambda}$ that has the highest posterior probability $p(\lambda|Y)$. Our forecasts and impulse response functions will be based on

$$p(\Phi, \Sigma_u, \theta | Y, \hat{\lambda}) = p(\Phi, \Sigma_u | Y, \theta, \hat{\lambda}) p(\theta | Y, \hat{\lambda}). \tag{23}$$

Thus, the joint posterior is given by the product of the marginal posterior of the DSGE model parameters and the conditional posterior of the VAR parameters given θ. Further details are provided in the Appendix.

3.2 Discussion

The prior mean $\Gamma_{xx}^{-1}(\theta)\Gamma_{yx}'(\theta)$ traces out a subspace of the VAR parameter space. For each value of θ the prior mean of Φ corresponds to the set of VAR(p) coefficients, that would minimize the expected one-step ahead quadratic prediction error loss if data were generated from the state-space representation of the DSGE model. Now suppose the DSGE model has an exact VAR(p) representation. If the hyperparameter λ is set to infinity, then the prior of Φ conditional on θ is a point mass at $\Gamma_{xx}^{-1}(\theta)\Gamma_{yx}'(\theta)$ and the VAR estimation would be equivalent to the direct estimation of the DSGE model. However, DSGE models by themselves often forecast poorly because they are tightly parameterized and impose inadequate cross-parameter restrictions on the vector autoregressive representation of the data. Therefore, it is desirable to assign some prior probability mass outside of the subspace traced out by $\Gamma_{xx}^{-1}(\theta)\Gamma_{yx}'(\theta)$, which is achieved by setting λ to a finite value. The orientation of the prior contours is such that the prior is fairly diffuse in the directions of the parameter space that we expect to estimate imprecisely according to the DSGE model.

DeJong, Ingram, and Whiteman (DIW, 1993) constructed a DSGE model prior as follows. They generated draws from the prior $p(\theta)$. For each $\theta^{(s)}$ draw, a long sequence of observations y_t^* was simulated from the DSGE model. They fitted a VAR to the simulated observations and obtained point estimates $\hat{\Phi}(\theta^{(s)}$ that are close to $\Gamma_{xx}^{-1}(\theta^{(s)})\Gamma_{yx}'(\theta^{(s)})$ and covariance estimates $\hat{V}(\theta^{(s)})$ for the VAR parameters. In order to obtain a conjugate prior for the VAR, DIW averaged $\hat{\Phi}(\theta^{(s)}$ and $\hat{V}(\theta^{(s)})$ across the $\theta^{(s)}$ draws, effectively eliminating the DSGE model parameters from the empirical model specification. While the DIW approach does not update the information with respect to the structural parameters θ, our procedure generates a joint posterior distribution for Φ, Σ_u, and θ that also allows posterior inference with respect to the DSGE model.

3.3 Non-Stationary Endogenous Variables

According to the DSGE model in Section 3 the log-levels of output and prices are non-stationary, which is consistent with the actual time series data. If the VAR is specified in levels then the asymptotic behavior of the sample moments $Y^{*'}_{T^*}Y^*_{T^*}$, $X^{*'}_{T^*}X^*_{T^*}$, and $Y^{*'}_{T^*}X^*_{T^*}$ changes. To be more specific, we will analyze $Y^{*'}_{T^*}Y^*_{T^*}$.

Let $z_t = [z'_{1,t}, z'_{2,t}]'$. The vector of endogenous variables y^*_t can be expressed as $y^*_t = \Xi z_t$, where $\Xi = [I_{2\times 2}, 0_{2\times 2}]$. Instead of analyzing the sample moments of y^*_t directly, we will examine the behavior of $\sum_{t=1}^{T_*} D_{T_*} z_t z_t' D'_{T_*}$, where D_T rotates and standardizes the elements of z_t. An appropriate rotation is given by

$$D_{T_*} = \left[\begin{array}{cc} T_*^{-1/2} I_{2\times 2} & -T_*^{-1/2}\Lambda \\ 0 & D_{T_*}^{(22)} \end{array}\right], \quad D_{T_*}^{(22)} = \left[\begin{array}{cc} T_*^{-1} & -T_*^{-1}\gamma/\ln m^* \\ 0 & T_*^{-3/2} \end{array}\right],$$

such that

$$D_{T_*} z_t = \left[\begin{array}{c} T_*^{-1/2}(\mu_1 + C_1(L)\epsilon_t) \\ T_*^{-1}[1, -\gamma/\ln m^*] \sum_{\tau=0}^{t} C_2(L)\epsilon_\tau \\ T_*^{-3/2}(\ln m^*)t + T_*^{-3/2}[0,1]\sum_{\tau=0}^{t} C_2(L)\epsilon_\tau \end{array}\right]. \tag{24}$$

Under the log-linearized DSGE model

$$T_*^{-1/2} \sum_{\tau=0}^{\lfloor [sT] \rfloor} C_2(L)\epsilon_\tau \Longrightarrow W(s), \tag{25}$$

where $\Longrightarrow$ signifies convergence in distribution and $W(s)$ is a bivariate Brownian Motion with covariance $C(1)\Sigma_{\epsilon\epsilon}C(1)'$.

All elements of z_t are now properly standardized and $\sum_{t=1}^{T_*} D_{T_*} z_t z_t' D'_{T_*}$ will converge in distribution

$$\sum_{t=1}^{T_*} D_{T_*} z_t z_t' D'_{T_*} \Longrightarrow \widetilde{\Gamma_{zz}}. \tag{26}$$

Unlike in the stationary case, $\widetilde{\Gamma}_{zz}$ is stochastic. Some of its elements are functionals of the Brownian Motion $W(s)$. The prior can be constructed from the expected value $I\!E[\widetilde{\Gamma_{zz}}]$. The sample moment $Y^{*'}_{T^*}Y^*_{T^*}$ is replaced by $\Xi D_{\lambda T}^{-1} I\!E[\widetilde{\Gamma_{zz}}] D_{\lambda T}^{-1'} \Xi'$. Similar calculations can be carried out for $X^{*'}_{T^*}X^*_{T^*}$ and $Y^{*'}_{T^*}X^*_{T^*}$.

4 Empirical Application

This section discusses the results obtained when we apply the priors derived from the DSGE model discussed in section 2 on a bivariate VAR in real output growth and inflation.[4] The section will first focus on the forecasting performance of the DSGE priors augmented VAR model and then discuss identification.

4.1 Forecasting

A far as forecasting goes, the results in this section are by no means general: they are specific to this particular bi-variate VAR and to this particular DSGE model. Other VARs, coupled with priors from other DSGE models, will give different results in terms of forecasting performance. The point of the section is not to provide upper or lower bounds on the extent to which DSGE models are helpful in forecasting. Rather, the objective of this section is twofold. First of all, the section shows that DSGE models *can* help in forecasting - a point already made by Ingram and Whiteman (1994) - and that the improvement is in some dimensions significant. We extend Ingram and Whiteman's finding by showing that DSGE model priors can be helpful in forecasting not only real, but also nominal variables. Second, the results show that the innovations presented in this paper are not just methodological niceties, but are also crucial in terms of forecasting. In particular, the fact that the DGSE prior is centered endogenously leads to a substantial improvement in forecasting performance relative to the case in which the prior over the DSGE model parameters is taken as exogenous (as in Ingram and Whiteman).

Does the DSGE model prior help in forecasting? If so, how much? Table 1 addresses these questions. The table shows the root mean square errors (*rmse*'s)

[4] The data for real output growth come from the Bureau of Economic Analysis (Gross Domestic Product-SAAR, Billions Chained 1996$), and the data for inflation come from the Bureau of Labor Statistics (CPI-U: All Items, seasonally adjusted, 1982-84=100). The data are available at a quarterly frequence from 1947:I to 2001:III. The lag length in the VAR is four quarters.

for forecasts of real output growth and inflation from one through eight, twelve, and sixteen quarters ahead. Specifically, for each series, the first column shows the percentage *rmse* for the unrestricted VAR ($\lambda = 0$), for the VAR augmented with the DSGE prior with $\lambda = 8$, and the percentage improvement (or loss, if negative) in *rmse* from using the DSGE prior. The symbols $*$ ($**$) indicate that the Diebold and Mariano (1995) asymptotic test rejects the null of no difference in mean square errors versus the alternative that the mean square error of the model with no prior is larger than that of the model with prior at the 10% (5%) level.[5] The symbols $\star$-$\star\star$, and †-††, have the same interpretation for the asymptotic test applied to the mean absolute error, and for the Diebold-Mariano sign test, respectively.
The table also shows the results for a multivariate forecasting performance statistics proposed by Doan, Litterman, and Sims (1984) and computed as the absolute value of the natural logarithm of the determinant of the error covariance matrix of the forecasts. Finally, the last column of the table shows the multivariate equivalent of the improvement in *rmse*'s.[6] The Diebold-Mariano tests are not applicable to the multivariate statistic.[7] The value $\lambda = 8$ is chosen to provide the most favorable result for the DSGE prior.

The table shows that in terms of multivariate statistics the DGSE model performs better than the unrestricted VAR for all forecast horizons. The improvement appears to be a monotonic function of the length of the forecast horizon: it is fairly trivial in the short run, but becomes more and more seizable as the forecast horizon

[5]We actually use Harvey, Leybourne, and Newbold (1997) modification of the Diebold-Mariano test. The results obtained with the original version of the test are not very different.

[6]This is computed by taking the difference between columns 8 and 7, divided by the number of variables (two), and again by two (to convert from variance to standard error) , and multiplied by 100 to obtain percentage figures. This number can be seen as the average in the improvements for the individual variables, adjusted to take into account the joint forecasting performance, i.e., the correlation in forecast errors.

[7]The differential in the multivariate statistic, unlike the differential in mean square error or mean absolute error, cannot be expressed as the average over time of loss differentials - condition which is necessary in order to apply the Diebold-Mariano test.

increases. For four years ahead forecasts the multivariate gain is above 5% - which in the forecasting literature is considered to be a fairly large number. Inspection of the forecasting performance of the individual series suggests that the multivariate improvement is mainly driven by gains in forecasting one series: inflation. The VAR with DSGE priors does marginally worse than the unrestricted VAR for one quarter ahead inflation forecasts, but it does substantially better in the medium and long run. For four years ahead forecasts of inflation the *rmse* for the VAR with DSGE priors is almost 9% lower than the *rmse* for the unrestricted VAR. Note that whenever the differences in *rmse* for inflation are large, they are also significant according to the Diebold-Mariano asymptotic test applied to both mean square error and mean absolute error, and to the Diebold-Mariano sign test.[8] In terms of output forecasts the VAR with priors does better than the unrestricted one up to two years ahead, but does worse in the longer run. The differences are fairly small, however, and never significant: the forecasting performance of the two models in terms of output is roughly the same for all horizons.

How does the VAR with DSGE priors perform relative to simple forecasting models, such as univariate models? Figure **??** shows the Theil-U statistics, that is, the ratio of the mean square error of the VAR versus that of the simple model, for two "simple models": a random walk and an AR(1) model. Specifically, the random walk model consists in a random walk in log-levels with drift for output, and in a random walk in log-differences without drift for CPI.[9] The latter model is considered fairly competitive in terms of forecasting inflation (see Atkeson ...). The AR(1) models are estimated in log-differences. Figure **??** shows that for one period ahead forecasts the univariate models outperform the VAR. For any forecast hori-

[8] The sign test is constructed under the assumption that the loss differentials (in our case, the differential in either square or absolute forecast errors) are i.i.d. Note that while the forecast errors are likely not i.i.d., especially at longer horizons, the forecast differential may well be (see Diebold and Mariano 1995, page 255, for a discussion of this point). As a matter of fact, we tested for serial correlation in forecast differentials and never found it significant.

[9] We chose the most competitive random walk model: a random walk model in log-differences for output, and a random walk in level for CPI do much worse in terms of forecasting.

zon beyond one quarter for output and two quarters for inflation, the VAR performs substantially better than the univariate models, especially at long forecast horizons. For eight quarters ahead the mean square error for the VAR is almost 25% that of the univariate models (implying a 50% improvement in terms of *rmse*'s). Figure **??** shows that the VAR with priors is competitive in terms of forecasting output at horizons beyond one quarter - in the sense that it outperforms simple models.

Next, we analyze the forecasting performance of the VAR, relative to the unrestricted VAR, as a function of the weight of the prior. Figure 2 plots the gain over the unrestricted VAR for real output growth, inflation, and the multivariate statistics. These figures correspond to those reported in columns 3, 6, and 9 of table 1. By definition, for $\lambda = 0$ the contour is equal to one. Figure 2 shows that the forecasting performance of the model is best in virtually all dimensions (output, inflation, and multivariate) for λ equal to 8, in the sense that all three surfaces show a ridge corresponding to this value of λ. For inflation the ridge is more pronounced for longer horizons, while for output it is more pronounced for shorter horizons. Overall, the surface is fairly flat for output, indicating that the forecasting performance does not vary dramatically as a function of the tightness of the prior. This is not the case for inflation: especially at longer forecast horizons the contour changes substantially as λ increases from 0 to 8. For larger values of λ the forecasting precision generally decreases, especially at long horizons, but the decline is very smooth. The result that the forecasting performance is maximized when the prior information coming from the DSGE model weights eight times the data is somewhat counterintuitive. The received wisdom is that DSGE models generally forecast poorly. As a consequence, one would think, the forecasting performance of the VAR should suffer if the weight of the prior is to high. The evidence presented here goes against this intuition: here the prior weights eight times the data in the computation of the posterior, and yet the forecasting performance is improved.

It is very interesting to compare the plots figure 2 with the equivalent plots obtained keeping constant the prior distribution of the DGSE model's parameters, shown in figure 3. When the distribution of DSGE parameters is taken as *exogenous* the forecasting performance of the VAR decreases very rapidly as λ increases, and is never superior to that of the unrestricted VAR. This result indicates that one of the methodological innovations of the paper, that is, to *endogenize* the distribution of the DSGE parameters via Bayes rule, is crucial in forecasting.

As mentioned in the introduction, whenever forecasters use prior information in VARs, they mostly use Minnesota priors. If we are to argue that forecasters should use a different kind of prior information, like that coming from a DSGE model, we better show that the new prior delivers better results, at least in some dimensions. This is precisely what we do in table 2. The table shows the improvement in terms of forecasting performance, relative to an unrestricted VAR, of three model: a VAR with DSGE priors only ($\lambda = 8$, $\iota = 0$), a VAR with Minnesota priors only ($\lambda = 0$, $\iota = 1.5$), and a VAR with both DSGE and Minnesota priors ($\lambda = 8$, $\iota = 1.5$). The symbols $*$-$**$, $\star$-$\star\star$, and †-††, have the same interpretation as in table 1.
The Minnesota prior is implemented as:

$$\bar{\phi} = (I_n \otimes (X_T' X_T + \lambda T \Gamma_{xx}) + \iota H_m^{-1})^{-1} (vec(X_T' Y_T + \lambda T \Gamma_{xy})) + \iota H_m^{-1} \phi_m) \tag{27}$$

where the parameter ι denotes the weight of the Minnesota prior (the value of 1.5 has been chosen to maximize the performance of the prior), ϕ_m is the prior mean and H_m is the prior tightness.[10] The values of ϕ_m and H_m are the same as in Doan, Litterman and Sims (1984), with the exception of the prior mean for the first lag. Since the Minnesota prior is applied to the VAR in growth rates - as opposed to log levels, to be consistent with the random walk hypothesis the prior mean for the first lag is zero and not one.[11] The numbers in columns 1, 4, and 7 of table 2 are the same ones reported in columns 3, 6, and 9 of table 1, and are repeated here for

[10] For reference, the value of ι in Doan, Litterman and Sims (1984) is 1.

[11] Consistently with Litterman's (1986) original implementation of the Minnesota prior, in choosing the tightness of the prior (i.e., the elements of the matrix H_m) for each equation 'own lags' are treated differently than 'other' lags. Because of this feature, the posterior distribution of the parameters does not have the convenient decomposition discussed in section 3. In addition, the

the sake of comparison.

The table shows that the VAR with Minnesota priors performs substantially better than the VAR with DSGE priors for output in the short run. The differences in forecasting performance are large and mostly significant. In the long run, however, its forecasting performance deteriorates notably relative to both the unrestricted and the DSGE-prior augmented VAR. In terms of inflation, the VAR with Minnesota priors performs always substantially worse than the VAR with DSGE priors. The VAR with *both* priors dominates the VAR with Minnesota priors only for all variables and all forecast horizons. Relative to the VAR with DSGE priors only, the VAR with both priors does substantially better for output in the short run (for one and two quarters ahead the improvement is greater than 5% and very significant) but tends to do worse as the forecast horizon increases. In terms of inflation forecasts, the VAR with DSGE priors only is always superior except for four years ahead forecasts.

Figure 4 provides an overview of the forecasting performance of the VAR model as a function of the weight of the DSGE (λ) and the Minnesota (ι) priors. The figure is interesting as it sheds some light on how the Minnesota prior and the DSGE prior interact. The figure displays on the z-axis the average improvement in the multivariate statistic over an unrestricted VAR ($\lambda = 0$,$\iota = 0$). The average is taken over the forecast horizon (from one to sixteen quarters ahead). The figure shows that the peak for the average improvement in multivariate forecasting performance is reached for $\lambda = 8$ and $\iota = 3$, with an average improvement over the unrestricted VAR of 2.8%. The average improvement for the VAR with DSGE prior only ($\lambda = 8$,$\iota = 0$) is just slightly lower than the peak (2.5%). The average improvement for the VAR with Minnesota priors only is much lower, and always below 1% (conditional on

value of ϕ corresponding to the mode of the posterior depends on Σ_u. To avoid these complications, whenever we also use DGSE prior we center the posterior estimates of the DSGE parameters without taking into account the presence of the Minnesota priors: we compute Γ_{xx} and Γ_{xy} as if ι were equal to zero, and then apply equation 27. If one wants to fully incorporate the Minnesota prior into the algorithm, one should use the version of the prior proposed by Kadiyala and Karlsson (1997).

$\lambda = 0$, the peak is reached for $\iota = 2.5$ with an average improvement of .6%).[12] The figure also shows that for given values of ι, the contour shows a ridge for λ equal to 8. For given λ, the shape of the contour with respect to ι is also fairly similar for different values of λ. This feature seems to indicate that the DSGE and Minnesota prior seem to operate independently from one another.

In summary, this section shows that the VAR with DSGE priors is a fairly competitive model in terms of forecasting. Especially for long run forecasts of inflation the model shows substantial improvement relative to an unrestricted VAR, a VAR with Minnesota prior, and simple univariate models. In terms of output forecast the VAR with DSGE priors is clearly inferior to univariate model for one quarter ahead forecasts, and is dominated by the Minnesota prior VAR in the short run, but otherwise the model holds its own. When the Minnesota priors are used in conjunction with the DSGE model prior short run output forecasts improve significantly, and outperform those of both the unrestricted VAR and the VAR with Minnesota priors only. The section also shows that centering the DSGE model parameters, that is, updating their prior distribution using information coming from the data, is crucial in terms of forecasting.

4.2 Identification

This section discusses the use of the DSGE model for the purpose of identification. The idea of using impulse responses form a DGSE model for the sake of identification is not new (Canova and De Nicolo (2001), Faust (1998), Uhlig (1999)). Rather, the contribution of this paper, in terms of identification, is twofold. The first contribution is that we use the same DSGE model for both forecasting and identification.

[12] In table 2 we chose to display the results for $\iota = 1.5$, as they magnify the differences in short run output performance with the unrestricted VAR, in spite of the fact that the average improvement is slightly lower (.5%). The VAR with ($\lambda = 0$,$\iota = 2.5$) does marginally better than the VAR with ($\lambda = 0$,$\iota = 1.5$) for medium run inflation forecasts - hence the slight improvement in multivariate forecasting performance - but its short run forecasting performance for output is not nearly as large.

There is an intuitive appeal for doing so. Although we do not want to formally propose the use of forecasting as a way to validate DSGE models, the fact that a prior from a DGSE model produces good forecasts means that the model itself fits the data reasonably well, in at least some dimensions. Hence, one might feel confident in using the model for identification.

The second contribution has to do with the fact that in out algorithm the distribution of DSGE parameters is updated so that it incorporates the information coming from the data. Generally, when researchers use DSGE models for identification, they follow these steps: i) choose a DSGE models, ii) choose a set of DSGE model parameters (sometimes choose a whole prior distribution over those parameters) via calibration , iii) rotate the impulse responses of the VAR so that they match those of the DSGE model. Note that in this procedure the data are used - aside from calibration - only in the estimation of the VAR. In our procedure there is an intermediate step between steps ii) and iii), which consists in updating the distribution of DSGE parameters.[13] What difference does the extra step make? The remainder of this section explores this question.

Figure 5 shows the impulse responses with respect to shocks to money supply and technology. The solid lines are the impulse responses corresponding to mode of the joint posterior distribution of the parameters. These are computed as follows. If θ^*, Σ_u^*, and ϕ^* are the set of parameters that maximize the joint posterior (computed according to expression 22), then the DSGE model delivers a set of impulse-responses corresponding to θ^*, and the VAR delivers a set of reduced-form impulse responses corresponding to Σ_u^* and ϕ^*. Using an orthonormal matrix we rotate the reduced-form impulse responses until they "match" the DSGE model impulse-responses. Specifically, we find the rotation that minimizes the average distance of the two sets of impulse-responses over a 40 period horizon.[14] The dashed lines represent the DSGE model impulse-responses (for θ equal to θ^*). The dashed-

[13] There are a few attempts in the literature to impose priors directly on impulse-responses (Dwyer 1998). Although this is not what we do here, our approach is close in spirit to this line of research.

[14] This is a modification of Canova and De Nicolo's (2000) procedure.

and-dotted lines represent the 90% error bands for impulse-responses. Again, the distribution of impulse-response functions is obtained by generating values of θ, Σ_u, and ϕ from the joint posterior distribution, and then applying the procedure just described.

Almost by definition, the VAR impulse responses are according to theory, at least in terms of their sign.[15] However, the magnitude and the shape of the VAR responses are quite different from the DSGE model's responses, especially as far as monetary shocks are concerned. In the model, the output response to a monetary shock is very small relative to the VAR's, and is not hump-shaped. Inflation's response to a monetary shock is very short lived in the model, while it is quite persistent in the VAR. Impulse responses to technology shocks are also quite different for the VAR and for the DSGE model.

Figure 6 addresses specifically the issue of endogenizing the DSGE model parameters, and the implications of this feature for identification. The thick solid and dashed-and-dotted lines in figure 6 are the same as in figure 5, and are reproduced here for convenience. The thin lines in figure 6 - solid and dashed-and-dotted - are obtained in the same way as the corresponding thick lines, except that the elements of θ, the vector of DSGE model parameters, are held fixed to some pre-specified values suggested by calibration.[16] The difference between the thick and the thin lines in figure 6 is precisely the difference made by the extra step - the updating of the distribution of the DGSE model parameters. Figure 6 shows that this difference can be very large. The only similarity between the two sets of impulse-responses is - by construction - in the sign. Otherwise the patterns look very dissimilar. The response of output with respect to a monetary shock is hump-shaped when the distribution of DSGE parameters is updated (thick line), but is not whenever the

[15] Note that even if the rotation procedure adopted here does not explicitly impose sign restrictions as in Canova and De Nicolo' (?), in practice the VAR impulse responses have the same sign as the model's

[16] Specifically, we use the following set of values: $\alpha = 0.35$, $\beta = 0.99$, $\gamma = 0.005$, $lnmst = 0.0002$, $\rho = 0.13$, $\psi = 0.65$, $\delta = 0.01$, $\sigma_\epsilon = 0.02$, $\sigma_\eta = 0.005$, $alp1rt = 50$.

DSGE model parameters are held fixed (thin line). Also, the relative importance of technology and money shocks for innovations in output is very different according to the two models: monetary shocks explain a larger fraction of short-run output fluctuations when the distribution of the DSGE parameters is updated than when it is help fixed. In many instances the 90% impulse-response bands do not even overlap, suggesting drastically different inferences from the two models. Some of the specific differences between the two sets of impulse responses depend of course on the model and on the choice of the values at which the DSGE model parameters are held fixed, so they cannot be generalized. Rather, the general implication from figure 6 is that updating the distribution of DSGE parameters makes a substantial difference for identification, and hence for policy analysis.

5 Conclusion

The paper takes the idea of Ingram and Whiteman (1994) - imposing priors from general equilibrium models on VARs - and develops it into a full-blown strategy, applicable to *any* log-linearized DSGE model and to *any* VAR, whether stationary or non-stationary, in reduced form or identified. The strategy involves the following steps: *i*) Choose a DSGE model and a prior distribution for its parameters (possibly via calibration); *ii*) Solve the linearized version of the DGSE model and map the prior distribution of parameters into a prior distribution for the VAR parameters; *iii*) Obtain via MonteCarlo methods the joint posterior distribution of DSGE and VAR parameters, which can then be used to compute predictive densities. Some of the steps - particularly *ii* and *iii* - may be demanding from the technical point of view, but the overall approach is very efficient in terms of computing time, as it does not involve computationally expensive simulations from the DSGE model (as in DeJong, Ingram, and Whiteman 1993). The latter feature is crucial in order to make the proposed strategy palatable to practitioners.

We apply the strategy to a VAR in inflation and output growth, and show that it is broadly successful in terms of forecasting performance. In no dimensions the VAR with DSGE priors does worse than the unrestricted VAR, and it does significantly better in a number of dimensions, particularly in terms of long-run forecasts of inflation. The VAR with DSGE priors is outperformed by simple forecasting models -like random walks and AR(1) - for very short run forecasts horizons (one quarter), but strikingly outperforms these models for any forecast horizon beyond two quarters. The VAR with Minnesota priors improves relative to the VAR with DSGE priors in terms of short and medium run output forecasts, but is not nearly as good in terms of forecasting inflation. When we use both DSGE and Minnesota priors, we obtain a model that does better than the VAR with only Minnesota priors in all dimensions. We also show that a key feature of the strategy proposed here - which is not present in Ingram and Whiteman (1994) - is the updating of the prior distribution of DSGE parameters: sticking to the parameters that come from the calibration exercise may lead to poor forecasting performance. Overall, these results confirm the finding from Ingram and Whiteman: priors from general equilibrium models can be of help in terms of forecasting.

Finally, we show that our approach can be used for the purpose of identification, and hence policy analysis. While the way to achieve identification - rotate the VAR impulse-responses until they conform the DSGE model's impulse-responses - is not new, our approach has the advantage of being embedded in the forecasting exercise. Instead of deriving the DSGE impulse-responses from an arbitrarily chosen DSGE model with arbitrarily chosen parameters, one can learn from the forecasting exercise which DSGE model and which parameterization performs better in terms of forecasting, and use it for the purpose of identification.

More research lies down the road. The paper shows how to address the issue of non-stationarity. However we discuss that issue only in passing, and use a stationary model in the empirical application. This is by choice. The issue of non-stationarity is both contentious and difficult. We thought that the discussion of non-stationarity might obscure the other points made in the paper. But we plan to investigate it in further work. On a different note, a feature of this paper is that the issue of identification is completely decoupled from that of forecasting: one could use any of the available approaches to identification in VARs and still use DSGE priors for the VAR parameters. This feature is both good and bad. Good as it makes the approach flexible, but bad as it contains a slight inconsistency. We use the DSGE model only as a "loose" prior for the VAR, but at the same time are very confident that the DSGE model's impulse-response functions have the "right" shape and sign. It would be nice to use the DSGE model as a prior for *both* the VAR parameters *and* identification. Again, we leave this task to future research.

References

Bernanke, Ben S. (1986): "Alternative Explanations of the Money-Income Correlation". In K.Brunner and A.Meltzer eds., *Real Business Cycles, Real Exchange Rates, and Actual Policies*, Carnegie-Rochester Series on Public Policy No. 25, Amsterdam: North-Holland, pp. 49-99.

Canova, Fabio and G. DeNicolo (2001): "Monetary Disturbances Matter for Business Cycle Fluctuations in the G-7". *Journal of Monetary Economics*, forthcoming.

Canova, Fabio (2001): "Validating Monetary DSGE Models through VARs". *Manuscript*, Universitat Pompeu Fabra.

Chang, Yongsung and Frank Schorfheide (2000): "Labor Supply Shifts and Economic Fluctuations". *PIER Working Paper*, **00-002**, University of Pennsylvania.

Christiano, Lawrence and Martin Eichenbaum (1992): "Liquidity Effects and the Monetary Transmission Mechanism". *American Economic Review*, **82**, 346-353.

Diebold, Francis X., and Roberto S. Mariano (1995): "Comparing Predictive Accuracy". *Journal of Business & Economic Statistics*, **13 (3)** , 253-263.

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

DeJong, David, Beth F. Ingram and Charles Whiteman (1993): "Analyzing VARs with Monetary Business Cycle Model Priors". *Proceedings of the American Statistical Association, Bayesian Statistics Section*, 160-169.

Dwyer, Mark (1998) "Impulse Response Priors for Discriminating Structural Vector Autoregressions." *Manuscript*, UCLA.

Faust, Jon (1998): "The Robustness of Identified VAR Conclusions about Money". *Carnegie Rochester Conference Series*, **49**, 207-244.

Harvey, David, Stephen Leybourne, and Paul Newbold (1994): "Testing the Equality of Prediction Mean Square Errors." *International Journal of Forecasting*, **13**, 281-291.

Ingram, Beth F. and Charles H. Whiteman (1994): "Supplanting the Minnesota prior – Forecasting macroeconomic time series using real business cycle model priors". *Journal of Monetary Economics*, **34**, 497-510.

Kadiyala, K. Rao, and Sune Karlsson (1997) 'Numerical Methods for the Estimation and Inference in Bayesian VAR-Models.' *Journal of Applied Econometrics*, **12**, pp. 99-132

Kim, Jinill (2000): "Constructing and Estimating a Realistic Optimizing Model of Monetary Policy". *Journal of Monetary Economics*, **45**, 329-360.

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

Litterman, Robert B. (1984): "Forecasting and Policy Analysis with Bayesian Vector Autoregression Models". *Federal Reserve Bank of Minneapolis Quarterly Review*, 30-41.

Litterman, Robert B. (1986): "Forecasting with Bayesian Vector Autoregressions – Five Years of Experience". *Journal of Business & and Economic Statistics*, **4**, 25-37.

Robert, Christian P. (1994): *"The Bayesian Choice"*. Springer-Verlag, New York.

Schorfheide, Frank (1998): "Econometric Modelling of Macroeconomic Aggregates", *Ph.D. Dissertation*, Yale University.

Schorfheide, Frank (2000): "Loss Function-Based Evaluation of DSGE Models". *Journal of Applied Econometrics*, **15**, 645-670.

Sims, Christopher A. (1986): "Are forecasting models usable for policy analysis?" *Quarterly Review of the Federal Reserve Bank of Minneapolis*, 10, pp. 2-16.

Sims, Christopher A. (1993): "A Nine-Variable Probabilistic Macroeconomic Forecasting Model", in: James H. Stock and Mark W. Watson (eds.) *"Business Cycles, Indicators, and Forecasting"*, NBER Studies in Business Cycles, Vol 28.

——— (1995): "Solving Rational Expectations Models". *Mimeographed*, Department of Economics, Yale University.

Sims, Christopher A. and Tao Zha (1998): "Bayesian methods for dynamic multivariate models". *International Economic Review*, **39 (4)**, pp. 949-968.

Todd, Robert M. (1984): "Improving Economic Forecasting with Bayesian Vector Autoregression". *Federal Reserve Bank of Minneapolis Quarterly Review*, 18-29.

Uhlig, Harald (1999): "What are the Effects of Monetary Policy? Evidence from an Agnostic Identification Procedure". *CEPR Working Paper.*

Zellner, Arnold (1971): *"Introduction to Bayesian Inference in Econometrics"*. John Wiley & Sons, New York.

A Computational Details

The model has the following hierarchical structure

$$p(Y,\Phi,\Sigma_u,\theta,\lambda) = p(Y|\Phi,\Sigma_u)p(\Phi,\Sigma_u|\theta,\lambda)p(\theta)p(\lambda). \tag{28}$$

The likelihood function $p(Y|\Phi,\Sigma_u)$ is given by

$$p(Y|\Phi,\Sigma_u) = (2\pi)^{-nT/2}|\Sigma_u|^{-T/2}\exp\left\{-\frac{1}{2}tr[\Sigma_u(Y-X\Phi)'(Y-X\Phi)]\right\}. \tag{29}$$

The prior density is based on the hypothetical pre-sample of observations from the DSGE model

$$\begin{aligned} p(\Phi,\Sigma_u|\theta,\lambda) &\propto p(Y^*(\theta,\lambda)|\Phi,\Sigma_u) \qquad (30)\\ &\propto (2\pi)^{-n\lambda T/2}|\Sigma_u|^{-\lambda T/2} \\ &\quad \times \exp\left\{-\frac{1}{2}tr[\lambda T\Sigma_u(\Gamma_{yy}-\Phi'\Gamma_{yx}'-\Gamma_{yx}\Phi+\Gamma_{xx})]\right\}. \end{aligned}$$

The matrices Γ_{yy}, Γ_{yx}, and Γ_{xx} are functions of the DSGE model parameters θ. The pre-sample likelihood function $p(Y^*(\theta,\lambda)|\Phi,\Sigma_u)$ does not integrate to one with respect to Φ,Σ_u. Thus, it does not provide a proper prior distribution for Φ,Σ_u. It integrates to

$$\begin{aligned} c(\theta) &= \int p(Y^*(\theta,\lambda)|\Phi,\Sigma_u)d(\Phi,\Sigma_u) \qquad (31)\\ &= (2\pi)^{-\frac{n(\lambda T-k)}{2}}|S^*|^{-\frac{\lambda T-k}{2}}|\lambda T\Gamma_{xx}|^{-\frac{n}{2}}2^{\frac{n(\lambda T-k)}{2}}\pi^{\frac{n(n-1)}{4}}\prod_{i=1}^{n}\Gamma[(\lambda T-k+1-i)/2], \end{aligned}$$

where $\Gamma[\cdot]$ denotes the gamma function and

$$S^* = \lambda T[\Gamma_{yy}-\Gamma_{yx}\Gamma_{xx}^{-1}\Gamma_{yx}'].$$

Thus, the joint density of data and parameters can be written as

$$p(Y,\Phi,\Sigma_u,\theta,\lambda) = p(Y|\Phi,\Sigma_u)p(Y^*(\theta,\lambda)|\Phi,\Sigma_u)c^{-1}(\theta,\lambda)p(\theta)p(\lambda). \tag{32}$$

In order to analyze the posterior distribution we use the following factorization

$$p(\Phi,\Sigma_u,\theta,\lambda|Y) = p(\Phi,\Sigma_u|Y,\theta,\lambda)p(\theta|Y,\lambda)p(\lambda|Y). \tag{33}$$

We assume that the parameter space of λ is finite $\Lambda = \{l_1, \dots, l_q\}$. Let $\pi_{j,0}$, $j = 1, \dots L$ be the prior probability of $\lambda = l_j$. The posterior odds ratio of $\lambda = l_i$ versus $\lambda = l_j$ is given by

$$\frac{\pi_{i,T}}{\pi_{j,T}} = \frac{\pi_{i,0} \int p(Y|\lambda = l_i)}{\pi_{j,0} \int p(Y|\lambda = l_j)}. \tag{34}$$

The marginal data density $p(Y|\lambda = l_i)$ is given by

$$\begin{aligned} p(Y|\lambda = l_i) &= \int p(Y|\Phi, \Sigma_u) p(Y^*(\theta, l_i)|\Phi, \Sigma_u) c^{-1}(\theta, l_i) p(\theta) d(\Phi, \Sigma_u, \theta) \qquad (35) \\ &= \int c^{-1}(\theta, \lambda_i) \left[\int p(Y|\Phi, \Sigma_u) p(Y^*(\theta, l_i)|\Phi, \Sigma_u) d(\Phi, \Sigma_u) \right] p(\theta) d\theta. \end{aligned}$$

Define $\tilde{p}(Y|\theta, \lambda)$ as

$$\begin{aligned} \tilde{p}(Y|\theta, \lambda) &= \int p(Y|\Phi, \Sigma_u) p(Y^*(\theta, l_i)|\Phi, \Sigma_u) d(\Phi, \Sigma_u) \qquad (36) \\ &= (2\pi)^{-\frac{n((\lambda+1)T-k)}{2}} |S|^{-\frac{(\lambda+1)T-k}{2}} |\lambda T \Gamma_{xx} + X'X|^{-\frac{n}{2}} \\ &\quad \times 2^{\frac{n((\lambda+1)T-k)}{2}} \pi^{\frac{n(n-1)}{4}} \prod_{i=1}^{n} \Gamma[((\lambda+1)T - k + 1 - i)/2], \end{aligned}$$

where

$$S = (\lambda T \Gamma_{yy} + Y'Y) - (\lambda T \Gamma_{yx} + Y'X)(\lambda T \Gamma_{xx} + X'X)^{-1} (\Gamma'_{yx} + X'Y).$$

Since expression (37) does not integrate to one with respect to Y it can be interpreted as a pseudo-marginal likelihood for θ. The actual marginal likelihood is given by

$$p(Y|\theta, \lambda) = c^{-1}(\theta, \lambda) \tilde{p}(Y|\theta, \lambda). \tag{37}$$

The posterior distribution of θ is given by

$$p(\theta|Y, \lambda) \propto c^{-1}(\theta, \lambda) \tilde{p}(Y|\theta, \lambda) p(\theta). \tag{38}$$

Let $\phi = vec(\Phi)$ and $\hat{\phi}$ be the OLS estimate of ϕ based on pre-sample and actual sample.

$$\hat{\phi} = vec((\lambda T \Gamma_{xx} + X'X)^{-1} (\lambda T \Gamma_{xy} + X'Y)) \tag{39}$$

The posterior distribution of ϕ and Σ_u conditional on θ and λ is

$$\begin{aligned} \Sigma_u|Y, \theta, \lambda &\sim \mathcal{IW}\Big(S, (1+\lambda) * T - k, n\Big), \qquad (40) \\ \phi|Y, \Sigma_u, \theta, \lambda, &\sim \mathcal{N}\Big(\bar{\phi}, \Sigma_u \otimes (\lambda T \Gamma_{xx} + X'X)^{-1}\Big), \qquad (41) \end{aligned}$$

where $\mathcal{IW}$ denotes the inverse Wishart distribution and $\mathcal{N}$ the multivariate normal distribution.

We use the following sampling scheme to generate draws from the joint posterior distribution of VAR parameters, DSGE model parameters, and the hyperparameter λ:

(i) For each $\lambda \in \Lambda$ use the Metropolis algorithm described in Schorfheide (2000) to generate draws from $p(\theta|Y, \lambda)$.

(ii) Based on these draws apply Geweke's (1999) modified harmonic mean estimator to obtain numerical approximations of the data densities $p(Y|\lambda)$.

(iii) Find the pre-sample size $\hat{\lambda}$ that has the highest data density.

(iv) Select the draws of $\{\theta_{(s)}\}$ that correspond to $\hat{\lambda}$ and use a standard method to generate draws from $p(\Phi, \Sigma_u|Y, \theta_{(s)}, \hat{\lambda})$ for each $\theta_{(s)}$.

Table 1: Forecasting performance - Unrestricted VAR versus VAR with DSGE priors

Horizon	Real GDP Growth			Inflation			Multivariate Stat		
	RMSE $\lambda = 0$	RMSE $\lambda = 8$	MSE Gain	RMSE $\lambda = 0$	RMSE $\lambda = 8$	MSE Gain	LnDet $\lambda = 0$	LnDet $\lambda = 8$	Gain
(Quarters)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	0.854	0.840	1.66 **	0.430	0.433	-0.55	20.505	20.526	0.53
2	1.333	1.316	1.31 *	0.843	0.832	1.24	18.246	18.295	1.22
4	2.037	2.015	1.06 †	1.672	1.642	1.78	16.034	16.089	1.38
6	2.354	2.327	1.15 †	2.734	2.676	2.12 ††	14.743	14.803	1.52
8	2.509	2.474	1.37 ††	4.082	3.959	3.02	13.797	13.874	1.95
10	2.575	2.550	0.98 †	5.446	5.220	4.16 **††	13.140	13.245	2.63
12	2.890	2.897	-0.22	6.815	6.427	5.69 ***††	12.481	12.620	3.46
14	3.311	3.356	-1.34	8.096	7.505	7.29 ***††	11.936	12.104	4.19
16	3.787	3.844	-1.51	9.351	8.521	8.87 ****††	11.532	11.744	5.32

Notes: All figures are in percent, except for columns 7 and 8. The symbols $*$ and $**$ ($\star$ and $\star\star$) indicate that the asymptotic Diebold-Mariano test rejects the null of no difference in mean square (absolute) errors at the 10% and 5% level, respectively. The symbols † and †† indicate that the Diebold-Mariano sign test rejects the null of no difference in forecasting performance at the 10% and 5% level, respectively. Period used to compute RMSE: 1977:II to 2001:III; total 82 quarters. The data used for estimation starts on 1952:II.

Table 2: Improvement in forecasting performance relative to an unrestricted VAR

Horizon	Real GDP Growth			Inflation			Multivariate Stat		
	$\lambda = 8$	$\lambda = 0$	$\lambda = 8$	$\lambda = 8$	$\lambda = 0$	$\lambda = 8$	$\lambda = 8$	$\lambda = 0$	$\lambda = 8$
	$\iota = 0$	$\iota = 1.5$	$\iota = 1.5$	$\iota = 0$	$\iota = 1.5$	$\iota = 1.5$	$\iota = 0$	$\iota = 1.5$	$\iota = 1.5$
(quarters)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	1.66 **	6.93 ***	7.17 ***†	-0.55	-1.76	-1.59	0.53	3.56	3.62
2	1.31 *	6.34 ****†	6.65 ****††	1.24	-0.52	0.64	1.22	3.67	4.30
4	1.06 †	5.08 ****	5.16 ****††	1.78	-1.42	-0.26	1.38	1.94	2.51
6	1.15 †	2.26 †	2.48 †	2.12 ††	-1.20	0.22	1.52	0.04	0.82
8	1.37 ††	-0.25	0.60 †	3.02	-0.29	1.79	1.95	-0.77	0.62
10	0.98 †	-2.27	-1.15	4.16 **††	0.55	3.40	2.63	-0.79	1.33
12	-0.22	-5.29	-4.24	5.69 ***††	1.44	5.42	3.46	-1.15	1.94
14	-1.34	-5.55	-4.88	7.29 ***††	2.80	7.83 *	4.19	-0.30	3.45
16	-1.51	-4.40	-3.85	8.87 ****††	4.25 *	10.32 **	5.32	1.11	5.62

Notes: All figures are in percent. λ and ι denote the weight of the DSGE and Minnesota priors, respectively. The symbols $\ast$ and $\ast\ast$ ($\star$ and $\star\star$) indicate that the asymptotic Diebold-Mariano test rejects the null of no difference in mean square (absolute) errors at the 10% and 5% level, respectively. The symbols † and †† indicate that the Diebold-Mariano sign test rejects the null of no difference in forecasting performance at the 10% and 5% level, respectively. Period used to compute RMSE: 1977:II to 2001:III; total 82 quarters. The data used for estimation starts on 1952:II.

Figure 1: Theil - U Statistics

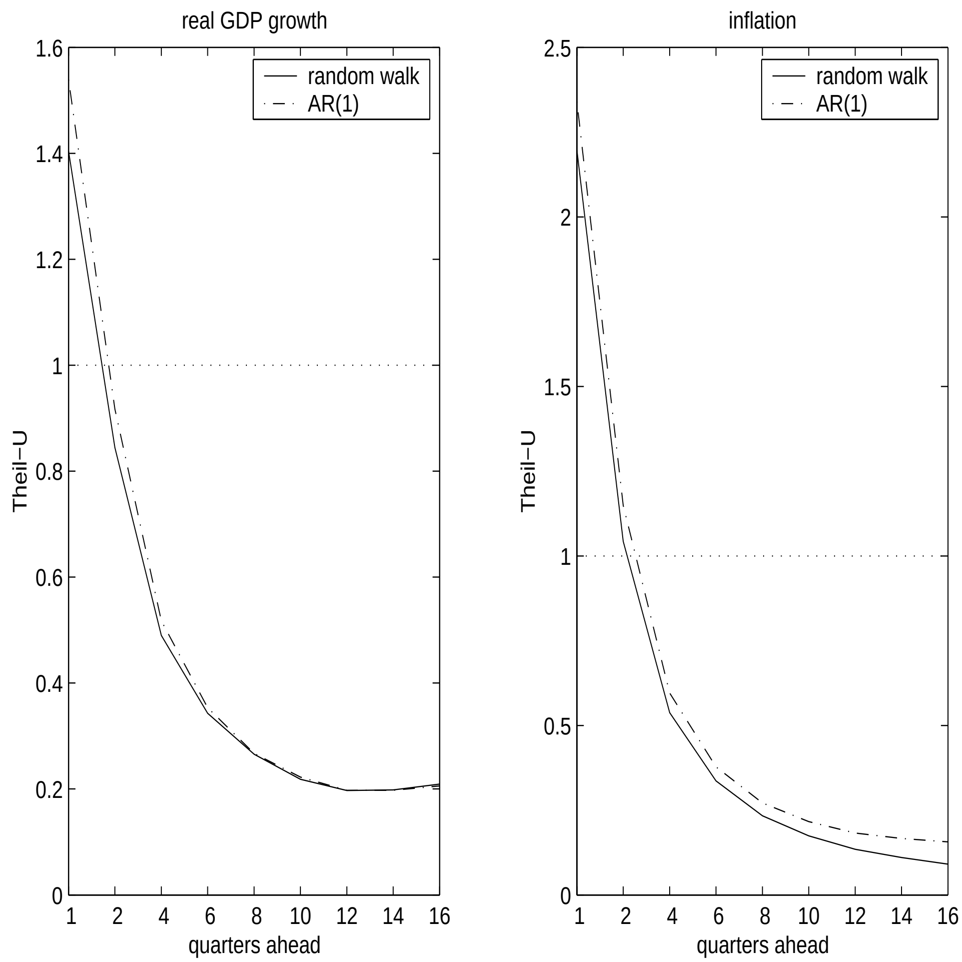
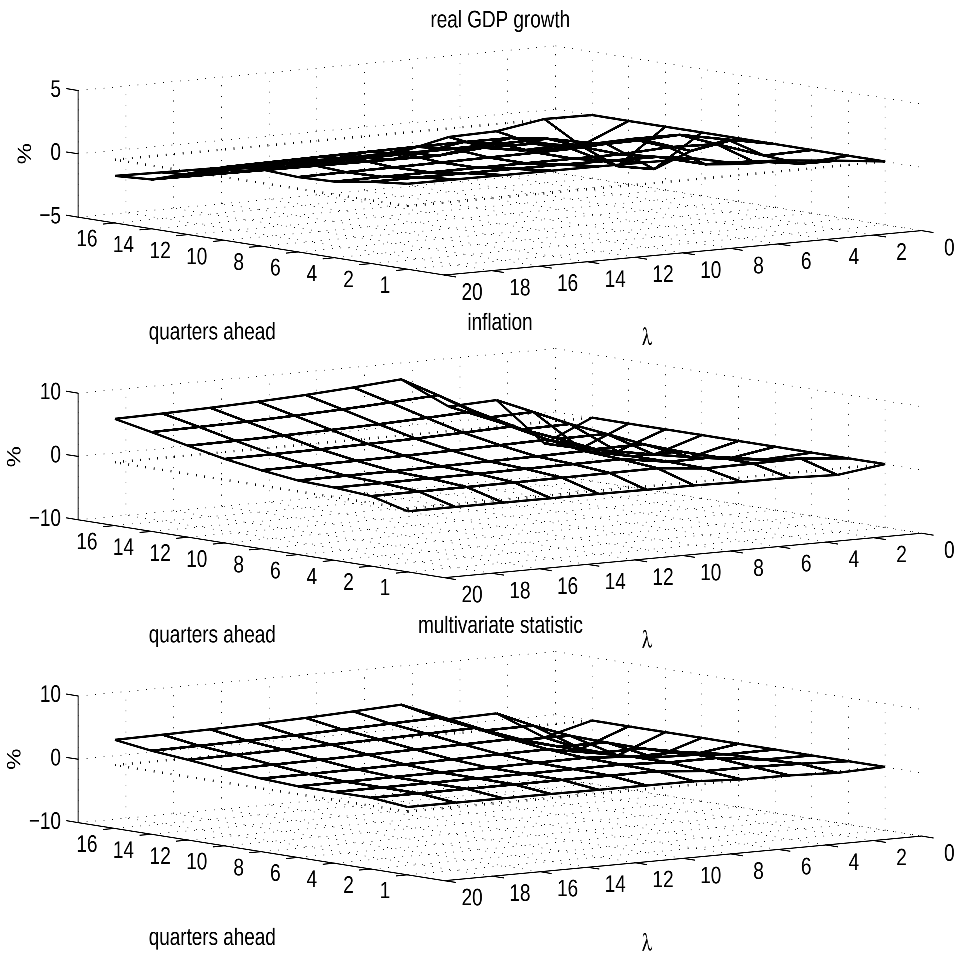
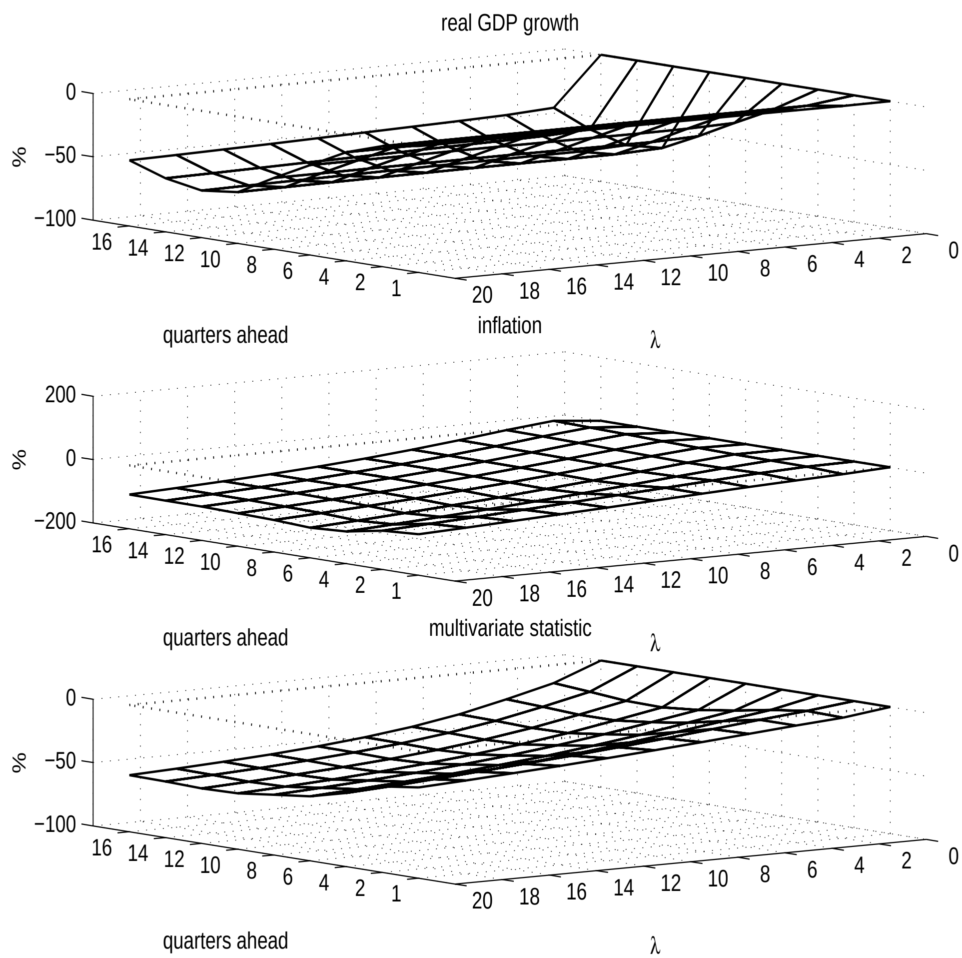


Figure 2: Forecasting performance as a function of the weight of the prior



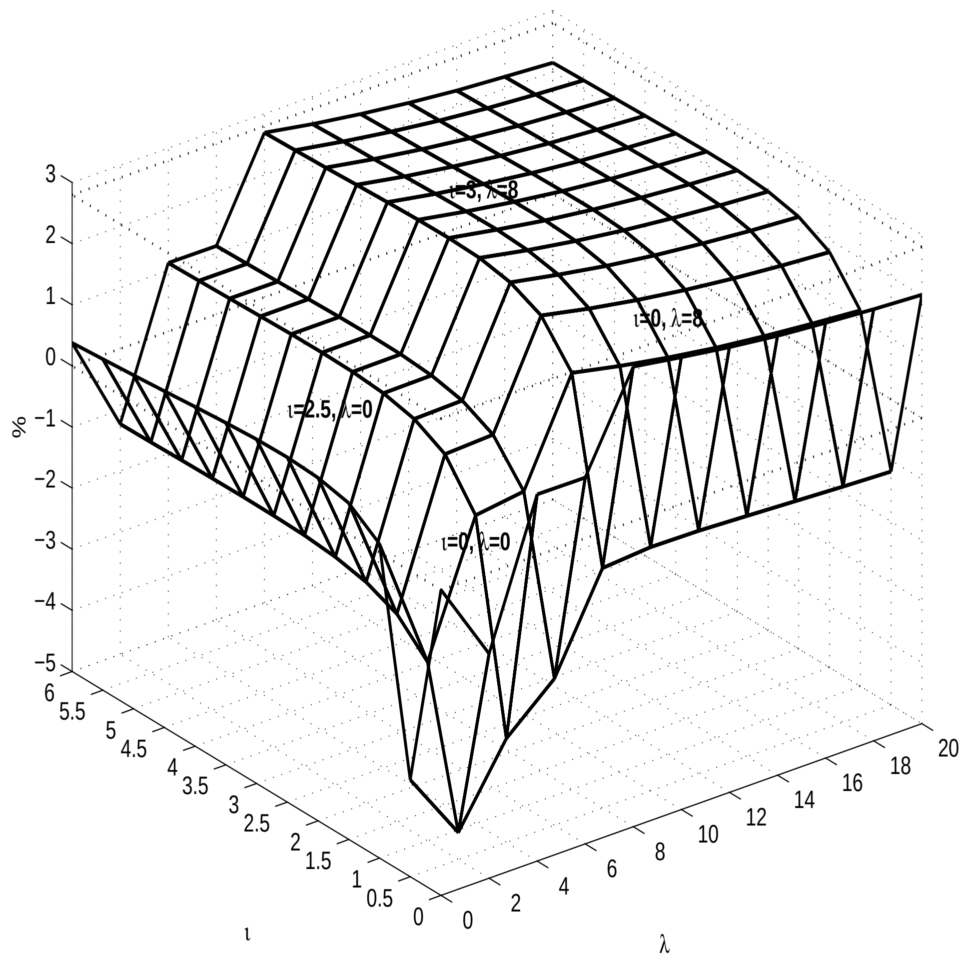
Note that, for the sake of visual clarity, the y-axis scale is reversed.

Figure 3: Forecasting performance as a function of the weight of the prior - exogenous DSGE paramters



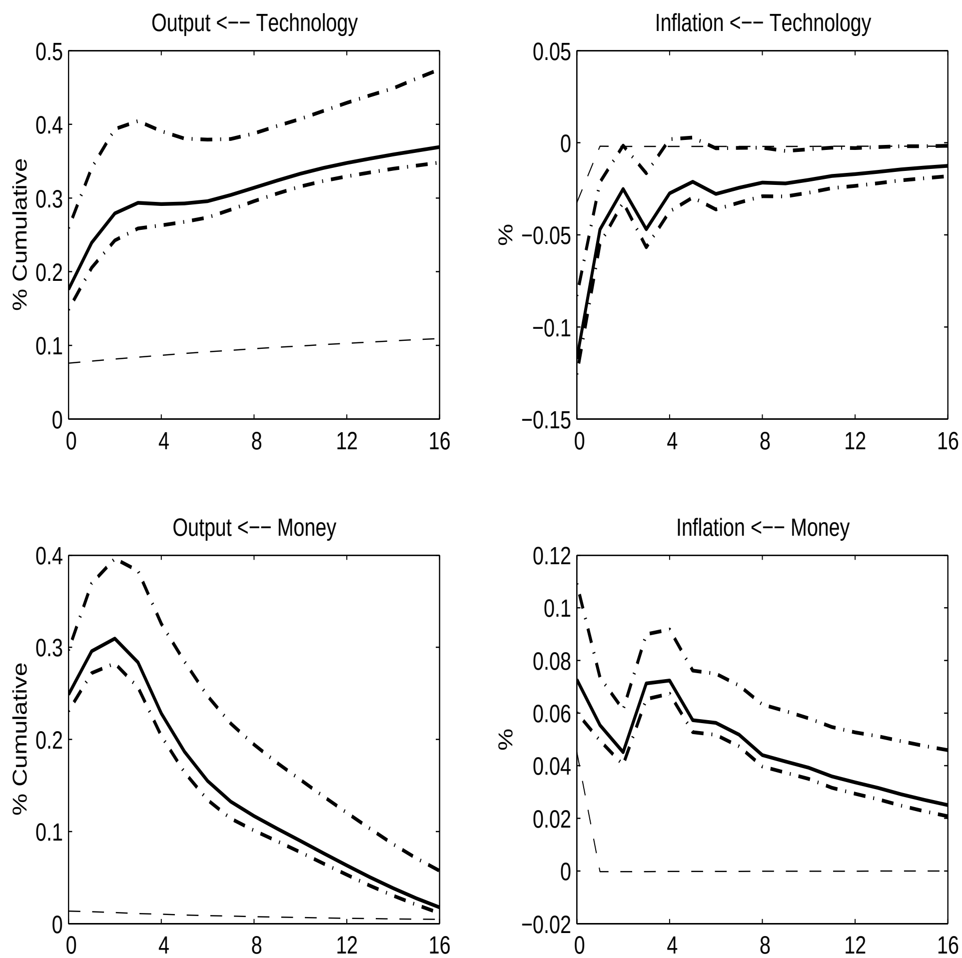
Note that, for the sake of visual clarity, the y-axis scale is reversed.

Figure 4: Average improvement in multivariate statistic as a function of the weight of the DSGE and Minnesota prior



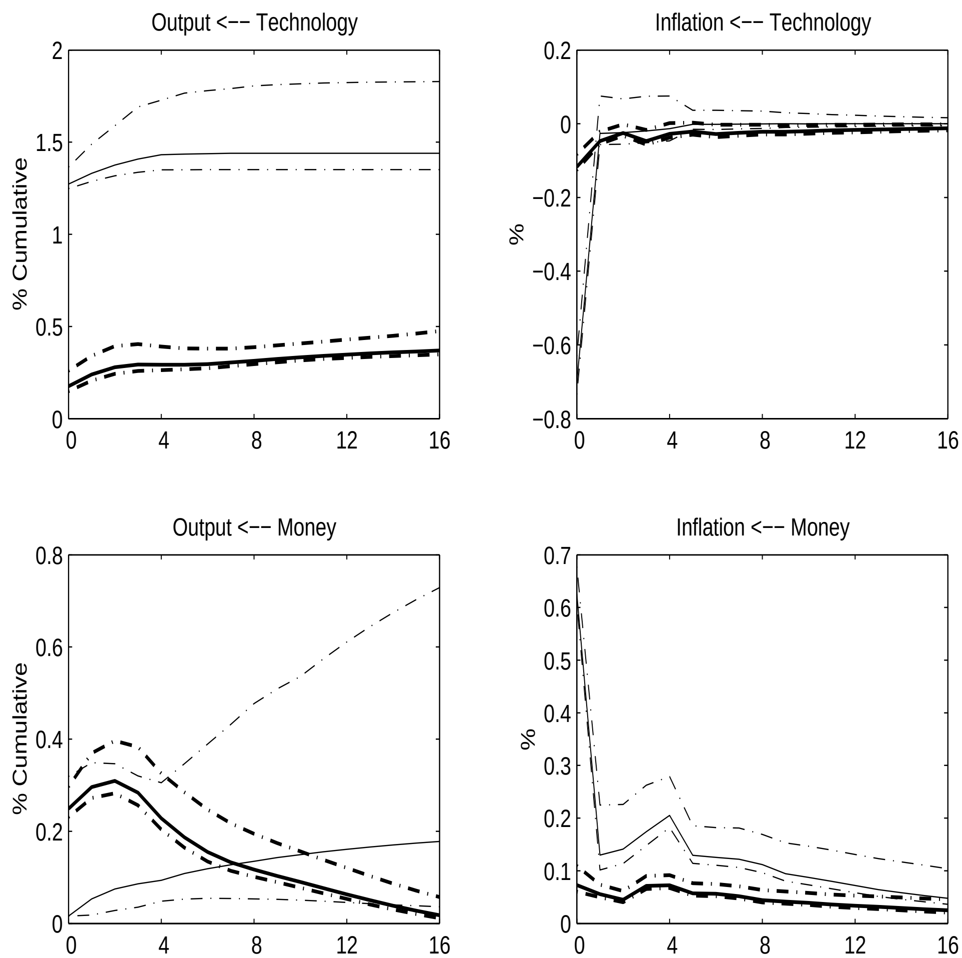
The plot shows the average improvement in the multivariate statistic over an unrestricted VAR ($\lambda = 0, \iota = 0$) as a function of the weight of the DSGE (λ) and the Minnesota (ι) priors. The average is taken over the forecast horizon (from one to sixteen quarters ahead).

Figure 5: Identified impulse response functions



The solid line represents the impulse responses corresponding to the mode of the posterior distribution. The dashed-and-dotted lines are 90% bands. The dashed line represents the impulse responses from the DSGE model.

Figure 6: Identified impulse response functions: endogenous vs exogenous DSGE model parameters



The thick solid lines represent the impulse responses corresponding to the mode of the posterior distribution. The thick dashed-and-dotted lines are 90% bands. The thin solid lines represent the impulse responses corresponding to a given set of DSGE model parameters. The thin dashed-and-dotted lines are 90% bands - also for given DSGE model parameters - obtained by drawing from the posterior distribution of VAR coefficients.