

Priors from General Equilibrium Models for VARs*

May 2003

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. We thank two anonymous referees, Sean Campbell, Fabio Canova, Frank Diebold, John Geweke, Dave Gordon, Thomas Lubik, Jim Nason, Chris Sims, Dan Waggoner, Charles Whiteman, Tao Zha, and seminar participants at the 2003 ASSA Meetings, Bank of Canada, Brown University, 2002 CIRANO Forecasting Conference, Duke University, European Central Bank, 2002 European Econometric Society Meetings, Federal Reserve Bank of Atlanta, 2002 LACEA Meetings, Spring 2002 Macro Federal Reserve System Meetings, Indiana University, 2002 Midwest Macro Conference, 2002 SED Meetings, University of Pennsylvania, and UCSD for helpful comments and discussions. We would like to thank Iskander Karibzhanov for translating some of our Matlab code in C. 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. Marco.DelNegro@atl.frb.org

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

Abstract

This paper uses a simple New-Keynesian dynamic stochastic general equilibrium model as a prior for a vector autoregression, shows that the resulting model is competitive with standard benchmarks in terms of forecasting, and can be used for policy analysis.

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models are popular nowadays in macroeconomics. They are taught in virtually every economics Ph.D. program, and represent a predominant share of publications in the field. Yet, when it comes to policy making, these models are scarcely used – at least from a quantitative point of view. The main quantitative workhorse for policy making at the Federal Reserve System is FRB-US, a macro-econometric model built in the Cowles foundation tradition – a style of macroeconomics that is no longer taught in many Ph.D. programs.¹ In their decision process, Fed policy makers rely heavily on forecasting. They want to know the expected path of inflation in the next few quarters, and by how much a 50 basis point increase in the federal funds rate would affect that path. FRB-US offers answers to these questions – answers that many macroeconomists would regard with suspicion given both the Lucas' (1976) critique and the fact that in general the restrictions imposed by Cowles foundation models are at odds with dynamic general equilibrium macroeconomics (Sims, 1980). General equilibrium models on the other hand have a hard time providing alternative answers. The fact that these models are perceived to do badly in terms of forecasting, as they are scarcely parameterized, is perhaps one of the reasons why they are not at the forefront of policy making.

While progress is being made in the development of DSGE models that deliver acceptable forecasts, e.g., Smets and Wouters (2003), this paper proposes an approach that combines a stylized general equilibrium model with a vector autoregression (VAR) to obtain a specification that both forecasts well and is usable for policy analysis. Specifically, the approach involves using prior information coming from a DSGE model in the estimation of a vector autoregression. Loosely speaking, this prior can be thought of as the result of the following exercise: (i) simulate time series data from the DSGE model, (ii) fit a VAR to these data. In practice we replace the sample moments of the simulated data by population moments computed from the DSGE model solution. Since the DSGE model depends on unknown structural

¹Some lucky few, among them the authors of this paper, have had the privilege of encountering proponents of this approach during their graduate studies.

parameters, we use a hierarchical prior in our analysis by placing a distribution on the DSGE model parameters. A tightness parameter controls the weight of the DSGE model prior relative to the weight of the actual sample. Markov-Chain Monte Carlo methods are used to generate draws from the joint posterior distribution of the VAR and DSGE model parameters.

The paper shows that the approach makes even a fairly stylized New Keynesian DSGE model competitive with standard benchmarks in terms of forecasting real output growth, inflation, and the nominal interest rate - the three variables that are of most interest to policy makers.² Up to this point our procedure borrows from the work of DeJong, Ingram, and Whiteman (1993), and Ingram and Whiteman (1994), who are the first to use priors from DSGE models for VARs. Ingram and Whiteman showed that prior information from the bare-bones stochastic growth model of King, Plosser, and Rebelo (1988) is helpful in forecasting real economic activity, such as output, consumption, investment, and hours worked.

In addition to documenting the forecasting performance of a trivariate VAR with a prior derived from a monetary DSGE model, this paper makes two contributions that significantly extend the earlier work. First, we show formally how posterior inference for the VAR parameters can be translated into posterior inference for DSGE model parameters. Second, we propose procedures to conduct two types of policy experiments within our framework. The first policy analysis is based on identified VAR impulse responses to monetary policy shocks. To obtain identification we construct an orthonormal matrix from the VAR approximation of the DSGE model to map the reduced form innovations into structural shocks. This procedure induces

² Ireland (2003) pursues a similar goal by augmenting a linearized DSGE model with unobservable errors, following Sargent (1989) and Altug (1989). Unlike in these earlier papers, Ireland does not restrict the errors to be uncorrelated across variables (“measurement errors”). Instead he allows them to follow a stationary VAR. While the resulting hybrid model is suitable for forecasting, Ireland provides no general discussion about the extent to which the latent errors are distinguishable from the structural shocks (identification) and no framework for policy analysis. Robertson, Tallman, and Whiteman (2002) also provide an alternative approach, based on relative entropy, to incorporate moment restrictions from economic theory into a model’s forecasts.

a DSGE model based prior distribution for the VAR impulse responses, which is updated with the sample information.

The second policy experiment is more ambitious. We want to forecast the effects of a change in the policy rule. Suppose a policy maker has a DSGE model that forecasts poorly and a VAR model that is silent about the effects of a policy regime change. In order to assess the effects of a policy intervention accurately, one has to correct the DSGE model predictions to bring its baseline forecasts under the existing policy regime on track. Based on the VAR with DSGE model prior (DSGE-VAR) we propose to implement such a correction. Since we lack knowledge about the relationship between policy parameters and DSGE model misspecification, we assume that the pre-intervention correction remains appropriate under the new policy regime. To illustrate our approach we try to forecast the impact of the change from the Martin-Burns-Miller regime to the Volcker-Greenspan regime on the volatility of inflation and compare it to the variability of actual inflation in the Volcker-Greenspan period.

The paper is organized as follows. Section 2 contains a brief description of the DSGE model that we are using to construct the prior distribution. Section 3 discusses the specification of the DSGE model prior and explores the joint posterior distribution of VAR and DSGE model parameters from a theoretical perspective. Empirical results for a VAR in output growth, inflation, and interest rates are presented in Section 4. Section 5 concludes. Proofs and computational details are provided in the Appendix.

2 A Simple Monetary DSGE Model

Our econometric procedure is applied to a trivariate VAR for output, inflation, and interest rates. The prior distribution for the VAR is derived from a New Keynesian DSGE model. To make this paper self-contained, we briefly review the model specification, which is adopted from Lubik and Schorfheide (2002). Detailed derivations

can be found, for instance, in King and Wolman (1999), King (2000), and Woodford (2003).

The model economy consists of a representative household, a continuum of monopolistically competitive firms, and a monetary policy authority that adjusts the nominal interest rate in response to deviations of inflation and output from their targets. The representative household derives utility from real balances M/P and consumptions C relative to a habit stock. We assume that the habit stock is given by the level of technology. This assumption assures that the economy evolves along a balanced growth path. The household derives disutility from hours worked, h , and maximizes

$$E_t \left[\sum_{s=t}^{\infty} \beta^{s-t} \left(\frac{(C_s/A_s)^{1-\tau} - 1}{1-\tau} + \chi \log \frac{M_s}{P_s} - h_s \right) \right], \quad (1)$$

where β is the discount factor, τ is the risk aversion parameter, and χ is a scale factor. P is the economy-wide nominal price level which the household takes as given. The (gross) inflation rate is defined as $\pi_t = P_t/P_{t-1}$.

The household supplies perfectly elastic labor services to the firms period by period and receives the real wage W . The household has access to a domestic capital market where nominal government bonds B are traded that pay (gross) interest R . Furthermore, it receives aggregate residual profits D from the firms and has to pay lump-sum taxes T . Thus, the households' budget constraint is of the form

$$C_t + \frac{B_t}{P_t} + \frac{M_t}{P_t} + \frac{T_t}{P_t} = W_t h_t + \frac{M_{t-1}}{P_t} + R_{t-1} \frac{B_{t-1}}{P_t} + D_t. \quad (2)$$

The usual transversality condition on asset accumulation rules out Ponzi-schemes.

The production sector is described by a continuum of monopolistically competitive firms each facing a downward-sloping demand curve for its differentiated product:

$$P_t(j) = \left(\frac{X_t(j)}{X_t} \right)^{-1/\nu} P_t. \quad (3)$$

This demand function can be derived in the usual way from Dixit-Stiglitz preferences, whereby $P_t(j)$ is the profit-maximizing price consistent with production level $X_t(j)$. The parameter ν is the elasticity of substitution between two differentiated

goods. The aggregate price level and aggregate demand X_t are beyond the control of the individual firm. Nominal rigidity is introduced by assuming that firms face quadratic adjustment costs in nominal prices. When a firm wants to change its price beyond the economy-wide inflation rate π^* , it incurs menu costs in the form of lost output:

$$\frac{\varphi}{2} \left(\frac{P_t(j)}{P_t(j-1)} - \pi^* \right)^2 X_t. \quad (4)$$

The parameter $\varphi \geq 0$ governs the degree of stickiness in the economy.

Production is assumed to be linear in labor $h_t(j)$, which each firm hires from the household:

$$X_t(j) = A_t h_t(j). \quad (5)$$

Total factor productivity A_t is an exogenously given unit-root process of the form

$$\ln A_t = \ln \gamma + \ln A_{t-1} + \tilde{z}_t \quad (6)$$

where

$$\tilde{z}_t = \rho_z \tilde{z}_{t-1} + \epsilon_{z,t} \quad (7)$$

and $\epsilon_{z,t}$ can be broadly interpreted as a technology shock that affects all firms in the same way. The specification of the technology process induces a stochastic trend into the model.³

Firm j chooses its labor input $h_t(j)$ and price $P_t(j)$ to maximize the present value of future profits:

$$\mathbb{E}_t \left[\sum_{s=t}^{\infty} Q_s D_s(j) \right] \quad (8)$$

subject to (5) and (6), where the time s profit is given by

$$D_s(j) = \left(\frac{P_s(j)}{P_s} X_s(j) - W_s h_s(j) - \frac{\varphi}{2} \left(\frac{P_s(j)}{P_{s-1}(j)} - \pi^* \right)^2 X_s(j) \right).$$

Here Q is the time-dependent discount factor that firms use to evaluate future profit streams. While firms are heterogenous ex ante, we only consider the symmetric equilibrium in which all firms behave identically and can be aggregated into

³Since our simple DSGE model lacks an internal propagation mechanism that can generate serially correlated output growth rates we assume that technology growth follows a stationary AR(1) process.

a single representative monopolistically competitive firm. Under the assumption that households have access to a complete set of state-contingent claims $Q_{t+1}/Q_t = \beta(C_t/C_{t+1})^\tau$ in equilibrium. Since the household is the recipient of the firms' residual payments it directs firms to make decisions based on the household's intertemporal rate of substitution.

The central bank follows a nominal interest rate rule by adjusting its instrument in response to deviations of inflation and output from their respective target levels:

$$\frac{R_t}{R^*} = \left(\frac{R_{t-1}}{R^*} \right)^{\rho_R} \left[\left(\frac{\pi_t}{\pi^*} \right)^{\psi_1} \left(\frac{X_t}{X_t^*} \right)^{\psi_2} \right]^{(1-\rho_R)} e^{\epsilon_{R,t}}, \quad (9)$$

where R^* is the steady state nominal rate and X_t^* is potential output, which we define as $X_t^* = A_t$ after normalizing the steady-state hours worked to one. The parameter $0 \leq \rho_R < 1$ determines the degree of interest rate smoothing. The central bank supplies the money demanded by the household to support the desired nominal interest rate. The monetary policy shock $\epsilon_{R,t}$ can be interpreted as an unanticipated deviation from the policy rule.

The government consumes a fraction ζ_t of each individual good j . We define $g_t = 1/(1 - \zeta_t)$ and assume that $\tilde{g}_t = \ln(g_t/g^*)$ follows a stationary AR(1) process

$$\tilde{g}_t = \rho_g \tilde{g}_{t-1} + \epsilon_{g,t}, \quad (10)$$

where $\epsilon_{g,t}$ can be broadly interpreted as government spending shock. The government levies a lump-sum tax (or subsidy) T_t/P_t to finance any shortfall in government revenues (or to rebate any surplus):

$$\zeta_t X_t + R_{t-1} \frac{B_{t-1}}{P_t} + \frac{M_{t-1}}{P_t} = \frac{T_t}{P_t} + \frac{M_t}{P_t} + \frac{B_t}{P_t} \quad (11)$$

The fiscal authority accommodates the monetary policy of the central bank and endogenously adjusts the primary surplus to changes in the government's outstanding liabilities.

To solve the model, optimality conditions are derived for the maximization problems. Consumption, output, wages, and the marginal utility of consumption are detrended by the total factor productivity A_t . The model has a deterministic steady

state in terms of the detrended variables. Define the percentage deviation of a variable y_t from its trend Y_t^* as $\tilde{y}_t = \ln Y_t - \ln Y_t^*$. The log-linearized system can be reduced to three equations in output, inflation, and nominal interest rates:

$$\tilde{x}_t = \mathbb{E}_t[\tilde{x}_{t+1}] - \tau^{-1}(\tilde{R}_t - \mathbb{E}_t[\tilde{\pi}_{t+1}]) + (1 - \rho_g)\tilde{g}_t + \rho_z \frac{1}{\tau} \tilde{z}_t, \quad (12)$$

$$\tilde{\pi}_t = \frac{\gamma}{r^*} \mathbb{E}_t[\tilde{\pi}_{t+1}] + \kappa[\tilde{x}_t - \tilde{g}_t], \quad (13)$$

$$\tilde{R}_t = \rho_R \tilde{R}_{t-1} + (1 - \rho_R)(\psi_1 \tilde{\pi}_t + \psi_2 \tilde{x}_t) + \epsilon_{R,t}, \quad (14)$$

where $r^* = \gamma/\beta$ is the steady state real interest rate and κ is a function of the price adjustment costs and the demand elasticity. The parameter measures the overall degree of distortion in the economy. Equation (12) is an intertemporal consumption Euler-equation, while (13) is derived from the firms' optimal price-setting problem and governs inflation dynamics. Equation (14) is the log-linearized monetary policy rule.⁴ The linear rational expectations system given by Equations (7), (10), and (12) to (14) can be solved, for instance, with the algorithm described in Sims (2002).

The relationship between the steady-state deviations and observable output growth, inflation, and interest rates is given by the following measurement equations:

$$\Delta \ln x_t = \ln \gamma + \Delta \tilde{x}_t + \tilde{z}_t \quad (15)$$

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

$$\ln R_t^a = 4[(\ln r^* + \ln \pi^*) + \tilde{R}_t].$$

In the subsequent empirical analysis a period t corresponds to one quarter. Output growth and inflation are quarter-to-quarter changes, whereas the interest rate, R_t^a , is annualized. The evolution of \tilde{x}_t , $\tilde{\pi}_t$, \tilde{R}_t , and \tilde{z}_t is given by the solution to the linear rational expectations system above. The DSGE model has three structural shocks which we collect in the vector $\epsilon_t = [\epsilon_{R,t}, \epsilon_{g,t}, \epsilon_{z,t}]'$. We assume that the

⁴In this paper we restrict the parameter space to values that lead to a unique stable solution of the linear rational expectations system. Lubik and Schorfheide (2002) discuss the econometric analysis of linear rational expectations models when the parameter space is not restricted to the determinacy region.

shocks are normally distributed and independent of each other and over time. Their standard deviations are denoted by σ_R , σ_g , and σ_z , respectively. The DSGE model parameters are stacked into the vector

$$\theta = [\ln \gamma, \ln \pi^*, \ln r^*, \kappa, \tau, \psi_1, \psi_2, \rho_R, \rho_g, \rho_z, \sigma_R, \sigma_g, \sigma_z]'. \quad (16)$$

The DSGE model imposes tight restrictions across the parameters of the moving-average (MA) representation for output growth, inflation, and interest rates.

3 Using the DSGE Model to Obtain a Prior for a VAR

Let $y_t = [\Delta \ln x_t, \Delta \ln P_t, \ln R_t^a]$. A less restrictive moving-average representation for the $n \times 1$ vector y_t than the one implied by the DSGE model of the previous section can be obtained from a vector autoregressive model:

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \dots + \Phi_p y_{t-p} + u_t, \quad (17)$$

where u_t is a vector of one-step-ahead forecast errors. VARs have a long tradition in applied macroeconomics as a tool for forecasting, policy analysis, and business cycle analysis. One drawback of VARs is that they are not very parsimonious: in many applications, data availability poses a serious constraint on the number of endogenous variables and the number of lags that can effectively be included in a VAR without overfitting the data. A solution to this problem of too many parameters is to constrain the parameter estimates by “shrinking” them toward a specific point in the parameter space.⁵ For instance, Doan, Litterman, and Sims (1984) proposed to shrink the estimates of a VAR that included the level of output, and prices, toward univariate random walk representations. The justification for this shrinkage was statistical rather than economic: no-change forecasts implied by the random-walk model were known to be fairly accurate. Our method shrinks the VAR estimates toward restrictions that are derived from a DSGE model. The spirit of our approach is to take the DSGE model restrictions seriously without imposing them dogmatically.

⁵Shrinkage estimators have a long tradition in statistics, see, e.g., Lehmann (1983).

Shrinkage estimators can be interpreted as Bayes estimators that are derived from prior distributions that concentrate much of their probability mass near the desired parameter restrictions. Priors are a way of systematically adding observations. In fact, many prior distributions can be incorporated into the estimation by augmenting the actual data set with dummy observations.⁶ Loosely speaking, we implement the DSGE model prior by generating dummy observations from the DSGE model, and adding them to the actual data. The ratio of dummy over actual observations – which we call λ in the remainder of the paper – measures the weight of the prior relative to the sample.

The DSGE model is indexed by a vector θ of so-called deep parameters. Instead of conditioning our analysis on a specific value of θ we place a prior distribution on the DSGE model parameters. In the econometric parlance, we construct a hierarchical prior consisting of a marginal distribution for θ and a conditional distribution for the VAR parameters given θ which is generated through the dummy observations. Bayes theorem then leads us to a joint posterior distribution for the DSGE model and VAR parameters. Our implementation of the DSGE model prior has an important advantage over previous approaches. We learn about the DSGE model parameters by implicitly searching for values of θ for which the distance between the VAR estimate and the vector autoregressive representation of the DSGE model is small. This information about θ will be exploited for the policy analysis in Section 4.

We have just given an informal description of our procedure. Much of the remainder of this section – for the econometrically oriented reader – provides an accurate and detailed exposition. In Section 3.1 we discuss the likelihood function. Section 3.2 describes the prior distribution. The posterior distribution is analyzed in Section 3.3.

⁶This so-called mixed estimation is originally due to Theil and Goldberger (1961), and used for instance, by Sims and Zha (1998).

3.1 Likelihood Function

In order to construct a likelihood function we assume that the innovations u_t in Equation (17) have a multivariate normal distribution $\mathcal{N}(0, \Sigma_u)$ conditional on past observations of y_t . Let Y be the $T \times n$ matrix with rows y'_t . Let $k = 1 + np$, X be the $T \times k$ matrix with rows $x'_t = [1, y'_{t-1}, \dots, y'_{t-p}]$, U be the $T \times n$ matrix with rows u'_t , and $\Phi = [\Phi_0, \Phi_1, \dots, \Phi_p]'$. The VAR can be expressed as $Y = X\Phi + U$ with likelihood function

$$\begin{aligned} p(Y|\Phi, \Sigma_u) &\propto \\ |\Sigma_u|^{-T/2} \exp \left\{ -\frac{1}{2} \text{tr}[\Sigma_u^{-1}(Y'Y - \Phi'X'Y - Y'X\Phi + \Phi'X'X\Phi)] \right\} \end{aligned} \quad (18)$$

conditional on observations y_{1-p}, \dots, y_0 . Although the DSGE model presented in Section 2 does not have a finite-order vector autoregressive representation in terms of y_t , the VAR can be interpreted as an approximation to the moving-average representation of the DSGE model. The magnitude of the discrepancy becomes smaller the more lags are included in the VAR. Since θ is of much lower dimension than the VAR parameter vector, the DSGE model imposes a restrictions on the (approximate) vector autoregressive representation of y_t .

3.2 Prior Distribution

Suppose the actual observations are augmented with $T^* = \lambda T$ artificial observations (Y^*, X^*) generated from the DSGE model based on the parameter vector θ . The likelihood function for the combined sample of artificial and actual observations is obtained by premultiplying (18) with

$$\begin{aligned} p(Y^*(\theta)|\Phi, \Sigma_u) &\propto \\ |\Sigma_u|^{-\lambda T/2} \exp \left\{ -\frac{1}{2} \text{tr}[\Sigma_u^{-1}(Y^{*\prime}Y^* - \Phi'X^{*\prime}Y^* - Y^{*\prime}X^*\Phi + \Phi'X^{*\prime}X^*\Phi)] \right\}. \end{aligned} \quad (19)$$

The factorization

$$p(Y^*(\theta), Y|\Phi, \Sigma_u) = p(Y^*(\theta)|\Phi, \Sigma_u)p(Y|\Phi, \Sigma_u) \quad (20)$$

suggests that the term $p(Y^*(\theta)|\Phi, \Sigma_u)$ can be interpreted as a prior density for Φ and Σ_u . It summarizes the information about the VAR parameters contained in the sample of artificial observations.

If we in fact would construct our prior by generating random draws from the DSGE model, a repeated application of our procedure would lead to stochastic variation in the prior distribution which is undesirable. In order to remove the stochastic variation from $p(Y^*(\theta)|\Phi, \Sigma_u)$ we replace the non-standardized sample moments $Y^{*'}Y^*$, $Y^{*'}X^*$, and $X^{*'}X^*$ by their expected values. According to our DSGE model, the vector y_t is covariance stationary and the expected values of the sample moments are given by the (scaled) population moments $\lambda T \Gamma_{yy}^*(\theta)$, $\lambda T \Gamma_{yx}^*(\theta)$, and $\lambda T \Gamma_{xx}^*(\theta)$, where, for instance, $\Gamma_{yy}^*(\theta) = \mathbb{E}_\theta[y_t y_t']$. Since these population moments can be computed analytically, our procedure is very efficient from the computational standpoint. Formally, the use of population moments implies that we replace expression (19) with:

$$\begin{aligned} p(\Phi, \Sigma_u | \theta) &= c^{-1}(\theta) |\Sigma_u|^{-\frac{\lambda T + n + 1}{2}} \\ &\times \exp \left\{ -\frac{1}{2} \text{tr} [\lambda T \Sigma_u^{-1} (\Gamma_{yy}^*(\theta) - \Phi' \Gamma_{xy}^*(\theta) - \Gamma_{yx}^*(\theta) \Phi + \Phi' \Gamma_{xx}^*(\theta) \Phi)] \right\}, \end{aligned} \quad (21)$$

where we also added an initial improper prior $p(\Phi, \Sigma_u) \propto |\Sigma_u|^{-(n+1)/2}$.

Provided that $\lambda T \geq k+n$ and $\Gamma_{xx}^*(\theta)$ is invertible, the prior density is proper and non-degenerate. In this case the normalization factor $c(\theta)$ can be chosen to ensure that the density integrates to one (the formula for $c(\theta)$ is given in the Appendix). Define the functions

$$\Phi^*(\theta) = \Gamma_{xx}^{*-1}(\theta) \Gamma_{by}^*(\theta) \quad (22)$$

$$\Sigma_u^*(\theta) = \Gamma_{yy}^*(\theta) - \Gamma_{nyx}^*(\theta) \Gamma_{xx}^{*-1}(\theta) \Gamma_{xy}^*(\theta). \quad (23)$$

Conditional on θ the prior distribution of the VAR parameters (21) is of the Inverted-Wishart (\mathcal{IW}) – Normal (\mathcal{N}) form⁷

$$\Sigma_u | \theta \sim \mathcal{IW}\left(\lambda T \Sigma_u^*(\theta), \lambda T - k, n\right) \quad (24)$$

⁷If $\Gamma_{xx}^*(\theta)$ is of reduced rank, then the prior concentrates its mass in a lower dimensional subspace of the domain of Φ . In our application the number of structural shocks equals the number of endogenous variables n to which the model is fitted and $\Gamma_{xx}(\theta)$ is full rank. For DSGE models

$$\Phi|\Sigma_u, \theta \sim \mathcal{N}\left(\Phi^*(\theta), \Sigma_u \otimes (\lambda T \Gamma_{xx}^*(\theta))^{-1}\right). \quad (25)$$

The specification of the prior is completed with a distribution of the DSGE model parameters, details of which we discuss in Section 4. Overall our prior has the hierarchical structure

$$p(\Phi, \Sigma_u, \theta) = p(\Phi, \Sigma_u | \theta)p(\theta). \quad (26)$$

The functions $\Phi^*(\theta)$ and $\Sigma_u^*(\theta)$ trace out a subspace of the VAR parameter space and can be interpreted as follows. Suppose that data are generated from a DSGE model with parameters θ . Among the p 'th order VARs the one with the coefficient matrix $\Phi^*(\theta)$ minimizes the one-step-ahead quadratic forecast error loss. The corresponding forecast error covariance matrix is given by $\Sigma_u^*(\theta)$.

Our prior is designed to assign probability mass outside of the subspace traced out by $\Phi^*(\theta)$ and $\Sigma_u^*(\theta)$.⁸ We use the covariance matrix $\Sigma_u^*(\theta) \otimes (\lambda T \Gamma_{xx}^*(\theta))^{-1}$ to distribute probability mass around $\Phi^*(\theta)$ and average over θ with respect to a prior $p(\theta)$. The orientation of the prior contours is such that the prior is fairly diffuse in the directions of the DSGE model parameter space that we expect to estimate imprecisely according to the DSGE model. Our prior differs from the one used by DeJong, Ingram, and Whiteman (1993), who used a simulation procedure to approximate (in our notation) the marginal prior for the VAR coefficients $p(\Phi, \Sigma) = \int p(\Phi, \Sigma_u | \theta)p(\theta)d\theta$ by a conjugate $\mathcal{IW} - \mathcal{N}$ prior.

3.3 Posterior Distribution

In order to study the posterior distribution we factorize it into the posterior density of the VAR parameters given the DSGE model parameters and the marginal posterior density of the DSGE model parameters:

$$p(\Phi, \Sigma_u, \theta | Y) = p(\Phi, \Sigma_u | Y, \theta)p(\theta | Y). \quad (27)$$

with less than n structural shocks we recommend to ensure that $\Gamma_{xx}^*(\theta)$ is full rank by introducing additional shocks or measurement errors.

⁸Ingram and Whiteman (1994) used a Gaussian prior for the DSGE model parameters $\theta \sim \mathcal{N}(\bar{\theta}, V_\theta)$ and approximated the function $\Phi^*(\theta)$ equation-by-equation with a first-order Taylor series around the prior mean $\bar{\theta}$ to induce a prior distribution for the VAR parameters.

Let $\tilde{\Phi}(\theta)$ and $\tilde{\Sigma}_u(\theta)$ be the maximum-likelihood estimates of Φ and Σ_u , respectively, based on artificial sample and actual sample

$$\tilde{\Phi}(\theta) = (\lambda T \Gamma_{xx}^*(\theta) + X'X)^{-1} (\lambda T \Gamma_{xy}^* + X'Y) \quad (28)$$

$$\begin{aligned} \tilde{\Sigma}_u(\theta) &= \frac{1}{(\lambda + 1)T} \left[(\lambda T \Gamma_{yy}^*(\theta) + Y'Y) \right. \\ &\quad \left. - (\lambda T \Gamma_{yx}^*(\theta) + Y'X)(\lambda T \Gamma_{xx}^*(\theta) + X'X)^{-1} (\lambda T \Gamma_{xy}^*(\theta) + X'Y) \right]. \end{aligned} \quad (29)$$

Since conditional on θ the DSGE model prior and the likelihood function are conjugate, it is straightforward to show, e.g., Zellner (1971), that the posterior distribution of Φ and Σ is also of the Inverted Wishart – Normal form:

$$\Sigma_u|Y, \theta \sim \mathcal{IW}\left((\lambda + 1)T\tilde{\Sigma}_u(\theta), (1 + \lambda)T - k, n\right) \quad (30)$$

$$\Phi|Y, \Sigma_u, \theta \sim \mathcal{N}\left(\tilde{\Phi}(\theta), \Sigma_u \otimes (\lambda T \Gamma_{xx}^*(\theta) + X'X)^{-1}\right). \quad (31)$$

The formula for the marginal posterior density of θ and the description of a Markov-Chain-Monte-Carlo algorithm that generates draws from the joint posterior of Φ , Σ_u , and θ are provided in the Appendix. The ability to compute the population moments $\Gamma_{yy}^*(\theta)$, $\Gamma_{xy}^*(\theta)$, and $\Gamma_{xx}^*(\theta)$ analytically from the log-linearized solution to the DSGE model and the use of conjugate priors for the VAR parameters makes the approach very efficient from a computational point of view: 25000 draws from the posterior distribution of all the items of interest - including forecast paths and impulse responses - can be obtained in less than 10 minutes using a 1.2GHz PC. In the remainder of this section we discuss the choice of λ and the source of information about the DSGE model parameters.

3.3.1 Choice of λ

The hyperparameter λ determines the effective sample size for the artificial observations, which is λT . If λ is small the prior is diffuse, and the actual observations dominate the artificial observations in the posterior. According to Equation (28), the posterior mean of Φ conditional on θ equals the OLS estimate of Φ if $\lambda = 0$.

For large values of λ the prior concentrates along the restriction functions $\Phi^*(\theta)$ and $\Sigma_u^*(\theta)$. We will show that as the length of the artificial sample increases the VAR parameter estimates will stay closer to the restrictions implied by the DSGE model. In the limit our procedure is equivalent to estimating the VAR subject to the DSGE model restrictions. Formally, as $\lambda \rightarrow \infty$, conditional on θ the posterior mean $\tilde{\Phi}(\theta)$ approaches $\Phi^*(\theta)$ and the variance $\tilde{\Sigma}_u(\theta)$ goes to zero. However, this result does not imply that the actual observations have no influence on the overall posterior distribution. An examination of the marginal posterior of the DSGE model parameters can clarify this point.

The posterior $p(\theta|Y)$ can be obtained by combining the marginal likelihood function

$$p(Y|\theta) = \int p(Y|\Phi, \Sigma_u) p(\Phi, \Sigma_u|\theta) d(\Phi, \Sigma_u) \quad (32)$$

with the prior $p(\theta)$ (see Appendix). If T is fixed and λ tends to infinity the marginal likelihood function $p(Y|\theta)$ approaches the (quasi)- likelihood function of the DSGE model⁹

$$p^*(Y|\theta) \propto |\Sigma_u^*(\theta)|^{-T/2} \exp \left\{ -\frac{1}{2} \text{tr} \left[\Sigma_u^{*-1}(\theta) (Y - X\Phi^*(\theta))' (Y - X\Phi^*(\theta)) \right] \right\}. \quad (33)$$

The function $p^*(Y|\theta)$ is obtained by replacing the unrestricted parameters Φ and Σ_u in Equation (18) with the restriction functions $\Phi^*(\theta)$ and $\Sigma_u^*(\theta)$. The result is summarized in the following proposition which is proved in the Appendix.

Proposition 1 *Let $\tilde{\theta}$ be the mode of $p(Y|\theta)$. For a fixed set of observations Y ,*

$$\ln \frac{p(Y|\theta)}{p(Y|\tilde{\theta})} \longrightarrow \ln \frac{p^*(Y|\theta)}{p^*(Y|\tilde{\theta})} \quad \text{as } \lambda \rightarrow \infty$$

uniformly for θ in compact subsets of Θ for which $\Sigma_u^(\theta)$ and $\Gamma_{xx}^*(\theta)$ are non-singular.*

Not surprisingly, the empirical performance of a VAR with DSGE model prior will crucially depend on the choice of λ . We use a data-driven procedure to determine

⁹Since the DSGE model typically does not have a finite-order vector autoregressive specification $p^*(Y|\theta)$ is a quasi-likelihood function from the perspective of the structural model.

an appropriate value $\hat{\lambda}$ of the hyperparameter. We maximize the marginal data density

$$p_\lambda(Y) = \int p_\lambda(Y|\theta)p(\theta)d\theta \quad (34)$$

with respect to λ over some grid $\Lambda = \{l_1, \dots, l_q\}$. Rather than averaging our conclusions about all possible values of λ , we condition on the value $\hat{\lambda}$ with the highest posterior probability. This marginal data density can also be used to choose an appropriate lag length for the VAR.

3.3.2 Learning about the DSGE Model Parameters

A major improvement of our procedure over earlier approaches is that it enables posterior inference with respect to the DSGE model parameters, whereas previous approaches only delivered a posterior for the VAR parameters. While the likelihood function (18) itself does not directly depend on the DSGE model parameters, the marginal distribution of θ will nevertheless be updated through the sample information. Similar to Smith (1993), an estimate of θ is implicitly obtained by projecting the VAR estimates onto the restrictions implied by the DSGE model.

The joint posterior density can be written as

$$p(\Phi, \Sigma_u, \theta|Y) = p(\Phi, \Sigma_u|Y)p(\theta|\Phi, \Sigma_u). \quad (35)$$

Learning about θ takes place indirectly through learning about the VAR parameters.

We expect that the best fit of the vector autoregression model is achieved for values of λ that allow moderate deviations from the DSGE model restrictions. Define the function

$$\begin{aligned} q(\theta|Y) &= \exp \left\{ -\frac{1}{2} \ln |\Sigma_u^{*-1}(\theta)| - \frac{1}{2} \text{tr}[\hat{\Sigma}_{u,mle}^{-1} \Sigma_u^*(\theta)] \right. \\ &\quad \left. - \frac{1}{2} \text{tr}[\hat{\Sigma}_{u,mle}^{-1} (\Phi^*(\theta) - \hat{\Phi}_{mle})' \Gamma_{xx}^*(\theta) (\Phi^*(\theta) - \hat{\Phi}_{mle})] \right\}. \end{aligned} \quad (36)$$

where $\hat{\Phi}_{mle}$ and $\hat{\Sigma}_{u,mle}$ maximize the likelihood function (18). Using a second-order Taylor expansion, it can be shown that the logarithm of $q(\theta|Y)$ is approximately a quadratic function of the discrepancy between the VAR estimates and the restriction

functions generated from the DSGE model. For long samples and under moderately tight priors the marginal log-likelihood function can be approximated as follows.

Proposition 2 *Let $\tilde{\theta}$ be the mode of $p(Y|\theta)$. Suppose $T \rightarrow \infty$, $\lambda \rightarrow 0$, and $\lambda T \rightarrow \infty$. Then*

$$\frac{1}{\lambda T} \ln \frac{p(Y|\theta)}{p(Y|\tilde{\theta})} = \ln \frac{q(Y|\theta)}{q(Y|\tilde{\theta})} + O_p(\max[(\lambda T)^{-1}, \lambda]).$$

The approximation holds uniformly for θ in compact subsets of Θ for which $\Sigma_u^(\theta)$ and $\Gamma_{xx}^*(\theta)$ are non-singular.*

The intuition for this result is the following. The weight of the prior relative to the likelihood function is small ($\lambda \rightarrow 0$), so that for all values of θ the posterior distribution of the VAR parameters concentrates around $\hat{\Phi}_{mle}$. The conditional density of θ given Φ and Σ_u projects $\hat{\Phi}_{mle}$ onto the subspace $\Phi^*(\theta)$. The amount of information accumulated in the marginal likelihood $p(Y|\theta)$ relative to the prior depends on the rate at which λT diverges. The more weight is placed on the artificial observations from the DSGE model (λ converges to zero slowly), the more curvature and information there is in $p(Y|\theta)$. According to Proposition 2 the posterior estimate of θ can be interpreted as a minimum distance estimate (e.g., Chamberlain (1984) and Moon and Schorfheide (2002)) that is obtained by minimizing the weighted discrepancy between the unrestricted VAR estimates $\hat{\Phi}_{mle}$ and the restriction function $\Phi^*(\theta)$.

4 Empirical Application

This section describes the results obtained when we apply the prior from the New Keynesian model described in Section 2 on a trivariate VAR in real output growth, inflation, and interest rates. We use quarterly data, and the lag length in the VAR is four quarters.¹⁰ Section 4 consists of four parts. Section 4.1 discusses the prior

¹⁰The 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

and posterior for the DSGE model parameters. Section 4.2 describes the forecasting results. We show that the DSGE model prior leads to a substantial improvement relative to an unrestricted VAR and in several dimensions dominates a VAR with Minnesota prior (Minn-VAR).

Policy analysis with a DSGE-VAR is discussed in Sections 4.3 and 4.4. First, we construct impulse response functions to study the effects of modest interventions (Leeper and Zha, 2003) in terms of unanticipated deviations from the monetary policy reaction function. The DSGE model is used to obtain an identification scheme for the VAR. Second, we use the DSGE-VAR to predict the effects of a policy rule change. While in the context of VARs the analysis of regime changes is generally subject to the Lucas' critique, our approach can be seen as a weighted average of two extremes: (i) using the DSGE model to forecast the effects of the policy change ($\lambda = \infty$), and (ii) using the VAR to make forecasts ($\lambda = 0$), thereby ignoring the effects of the policy intervention. In our framework, the choice of the prior weight λ reflects the degree of misspecification of the structural model. We try to predict the impact of the change from the Martin-Burns-Miller regime to the Volcker-Greenspan regime using the DSGE-VAR. The results suggest that the approach is promising, at least in some dimensions.

4.1 Prior and Posterior of θ

All empirical results are generated with the prior distribution reported in Table 1. The model parameters $\ln \gamma$, $\ln \pi^*$, $\ln r^*$, σ_R , σ_g , and σ_z are scaled by 100 to convert their units into percentages. The priors for the quarterly steady state growth rate, inflation rate, and real interest rate are fairly diffuse and have means 0.5%, 1.0%, and 0.5%, respectively. With 90% prior probability the risk aversion parameter τ is between 1.2 and 2.8, whereas the slope of the Phillips curve κ is between 0.06

Statistics (CPI-U: All Items, seasonally adjusted, 1982-84=100). The interest rate series are constructed as in Clarida, Gali, 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. Our sample ranges from 1955:III to 2001:III.

and 0.51. The latter interval is consistent with the values that have been used in calibration exercises, e.g., Clarida, Galí, and Gertler (2000). The priors for the policy parameters ψ_1 and ψ_2 are centered at Taylor's (1999) values.¹¹ The prior is truncated at the indeterminacy region of the parameter space.

As stressed in Section 3, our procedure also generates posterior estimates for the DSGE model parameters. Such estimates are presented in Table 2 for the sample period 1959:III to 1979:II. To illustrate that the extent of learning about θ depends on the weight λ of the DSGE model prior, Table 2 reports 90% posterior confidence sets for $\lambda = 1$ and $\lambda = 10$. A comparison of prior and posterior intervals indicates that for $\lambda = 1$ the data lead to a modest updating. The confidence intervals for most parameters shrink. For instance, the prior confidence interval for the interest rate smoothing parameter ρ_R ranges from 0.15 to 0.81, whereas the posterior interval narrows to [0.21, 0.54]. The parameter that characterizes the responsiveness of the central bank to inflation is revised downward, whereas the confidence interval for the output coefficient shifted upwards. The posterior means for ψ_1 and ψ_2 are 1.3 and 0.3, respectively. The updating is slightly more pronounced for $\lambda = 10$, when the artificial sample size is ten times as long as the actual sample.

4.2 Forecasting Results

The objective of this subsection is to show that VARs with DSGE model priors produce forecasts that improve on those obtained using unrestricted VARs, and are competitive with those obtained using the popular Minnesota prior. The Minnesota prior shrinks the VAR coefficients to univariate unit root representations. While it has been empirically successful, e.g., Litterman (1986), Todd (1984), it lacks economic justification and ignores information with respect to co-movements of the

¹¹Since the inflation rate and the interest rate in the DSGE model are quarter-to-quarter, the value of ψ_2 corresponds to one fourth of the value obtained in univariate Taylor-rule regressions that use annualized interest rate and inflation data.

endogenous variables.¹²

In a particular instance, this point has already been made by Ingram and White-
man (1994). However, we provide two extensions of their results. First, we show
that DSGE model priors can be helpful in forecasting not only real but also nominal
variables. Second, while Ingram and Whiteman document the ex post forecasting
performance of their DSGE-VAR as a function of the relative weight of the prior,
we also study the forecasting performance under the requirement that a hyperpa-
rameter $\hat{\lambda}$ is chosen *ex ante* for each forecast based on the marginal posterior data
density $p_\lambda(Y)$. This is an important extension because the forecasting performance
of the VAR is sensitive to λ and it has to be guaranteed that a good λ can be chosen
before the actual forecast errors become available.

Most of the remainder of the section will present results from a forecasting
exercise using a rolling sample from 1975:III to 1997:III (90 periods). For each date
in the forecasting interval we used 80 observations in order to estimate the VAR,
that is, a ratio of data to parameters of about 6 to 1 for each equation. This choice is
motivated by the fact that the data-parameter ratio in larger models that are being
used for actual forecasting, such as the Atlanta Fed VAR, is of the same magnitude.
It is important to remark that the results presented in this section have no pretense
of being general: they are specific to the particular DSGE model, and the particular
VAR being estimated.

¹² In this paper the Minnesota prior is implemented as:

$$\bar{\phi} = (\mathcal{I}_n \otimes (X'_T X_T) + \iota H_m^{-1})^{-1} (vec(X'_T Y_T) + \iota H_m^{-1} \phi_m)$$

where the parameter ι denotes the weight of the Minnesota prior, ϕ_m is the prior mean and H_m
is the prior tightness. 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 of output growth and inflation. Since
these two variables enter 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 of the ‘own’ regressor in the output
growth and inflation equations is zero and not one. The Minnesota prior is augmented by a proper
 \mathcal{IW} prior for Σ_u . The weight of the Minnesota prior is controlled by the hyperparameter ι . The
hyperparameter is selected *ex ante* using a modification of (34). This value hovers around 0.5,
depending on the sample. The value used in Doan, Litterman and Sims (1984) is $\iota = 1$.

We begin by investigating how the forecasting performance of the VAR changes as a function of λ and the forecast horizon. Figure 1 plots the percentage improvement (or loss, if negative) in root mean square forecast errors (*rmse*) of the DSGE-VAR relative to the unrestricted VAR for cumulative real output growth, cumulative inflation, and the federal funds rate.¹³ The fourth panel of Figure 1 depicts the improvements in the multivariate forecasting performance statistic proposed by Doan, Litterman, and Sims (1984).¹⁴ The grid of values of λ ranges from 0 to ∞ , where $\lambda = \infty$ means forecasting with the VAR approximation of the DSGE model.

By definition the gain for $\lambda = 0$ is zero. As the weight of the prior is increased we observe a substantial gain in the forecast performance. The surface for the multivariate statistic is fairly flat for values of λ between 0.5 and 5, but then deteriorates sharply as λ approaches infinity. Overall our performance measures have an inverse U-shape as a function of the hyperparameter λ . This indicates that there is a benefit from shrinking the VAR estimates toward the DSGE model restrictions without dogmatically imposing them. The ex post optimal λ for long-run forecasts tends to be larger than for short-run forecasts. In order to obtain accurate forecasts over long horizons one has to estimate powers of the autoregressive coefficients Φ . The large sampling variance of these estimates can be reduced by increasing the weight of the prior. However, once the length of the artificial sample relative to the actual sample exceeds 2, the variance reduction is dominated by an increased bias and the forecasting accuracy generally deteriorates. Interestingly, the deterioration is not sharp at all: in particular, for inflation and the interest rate the long-run forecasts from DSGE-VAR are still accurate even when the prior weight is infinity.

While Figure 1 documents that the DSGE model prior leads to an improved

¹³Neither the output growth rates nor the inflation rates are annualized.

¹⁴The *ln-det* statistic is defined as the converse of the natural logarithm of the determinant of the error covariance matrix of the forecasts, divided by two (to convert from variance to standard error) times the number of variables that are forecasted (to obtain an average figure). The improvement in the multivariate forecasting performance statistics is computed by taking the difference between the multivariate statistics 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.

forecasting performance, two questions remain. First, can an appropriate value of λ be selected *ex ante*? Second, how does the DSGE-VAR compare to a more competitive forecasting model such as a VAR with Minnesota prior?

We argued in Section 3.3.1 that λ can be chosen over a grid Λ to maximize the marginal data density $p_\lambda(Y)$, given in Equation (34). This leads to the hyperparameter estimate

$$\hat{\lambda} = \operatorname{argmax}_{\lambda \in \Lambda} p_\lambda(Y). \quad (37)$$

Depending on the sample, this value generally hovers around 0.6, which corresponds to 48 artificial observations from the DSGE model. However, the shape of the marginal data density as a function of λ is flat for values of λ between 0.5 and 2, suggesting that the fit of the model is approximately the same within that range. These estimates are broadly consistent with the value that leads to the best *ex-post* one-step-ahead forecast performance.¹⁵

Table 3 documents the *rmse* improvements of the DSGE-VAR based on $\hat{\lambda}$ relative to the unrestricted VAR and the Minn-VAR. The DSGE-VAR clearly outperforms the unrestricted VAR, even if the weight of the prior is chosen *ex ante*. The DSGE-VAR also dominates the Minn-VAR in terms of output growth and inflation forecasts and according to the multivariate performance measure. While the improvement is small, around 1 to 1.5 percent, for short-horizons, it is substantial for forecast horizons beyond 2 years. The DSGE-VAR forecasts of the federal funds rate are unfortunately slightly worse than the Minn-VAR forecasts for most of the horizons that we considered. Overall, these results suggest that the DSGE-VAR is competitive with, and to some extent improves upon, the Minn-VAR.¹⁶ When

¹⁵A full Bayesian procedure would average over λ rather than condition on the highest posterior probability λ . However, in our experience the values of λ that have non-negligible posterior probability produce very similar predictions so that the gain from averaging instead of conditioning is minimal.

¹⁶We do not report formal significance tests for superior forecast performance, such as the Diebold and Mariano (1995) test, since the assumptions underlying those tests do not match the setup in our paper. Thus, the results should be interpreted as *ex post* accuracy comparisons, not as hypothesis tests. Although not pursued here, Bayesian posterior odds could be used to choose among DSGE-

interpreting these results one has to bear in mind a key difference between Minnesota and DSGE model prior. The Minnesota prior has only a statistical but not an economic justification. We will show subsequently that the DSGE model prior can be exploited in the analysis of policy interventions.

4.3 Impulse Response Functions

In order to compute dynamic responses of output, inflation, and interest rates to unanticipated changes in monetary policy and to other structural shocks it is necessary to determine the mapping between the structural shocks ϵ_t and the one-step-ahead forecast errors u_t . Let Σ_{tr} be the Cholesky decomposition of Σ_u . It is well known that in any exactly identified structural VAR the relationship between u_t and ϵ_t can be characterized as follows:

$$u_t = \Sigma_{tr}\Omega\epsilon_t, \quad (38)$$

where Ω is an orthonormal matrix and the structural shocks are from now on standardized to have unit variance, that is $\mathbb{E}[\epsilon_t\epsilon_t'] = \mathcal{I}$. According to Equation (17) the initial impact of ϵ_t on the endogenous variables y_t in the VAR is given by

$$\left(\frac{\partial y_t}{\partial \epsilon_t'} \right)_{VAR} = \Sigma_{tr}\Omega. \quad (39)$$

The identification problem arises from the fact that the data are silent about the choice of the rotation matrix Ω . More prosaically, since $\Sigma_{tr}\Omega\Omega'\Sigma_{tr}' = \Sigma_{tr}\Sigma_{tr}' = \Sigma_u$ the likelihood function is invariant to Ω .

Macroeconomists generally require Ω to have some ex ante justification and to produce ex post impulse response functions that are “reasonable”, i.e., conform in one or more dimensions with the predictions of theoretical models. Since there is no agreement on what these dimensions should be, a multitude of identification strategies have been proposed (see Christiano, Eichenbaum, and Evans (1999) for a recent survey). In this paper we have used the DSGE model to derive a prior distribution for the reduced form VAR parameters. Hence, it is quite natural in our

VAR and Minn-VAR *ex ante*.

framework to use the structural model also to identify the VAR. Thus, we will now construct a rotation matrix Ω based on the dynamic equilibrium model.

The DSGE model itself is identified in the sense that for each value of θ there is a unique matrix $A_0(\theta)$, obtained from the state space representation (15), that determines the contemporaneous effect of ϵ_t on y_t . Using a QR factorization of $A_0(\theta)$, the initial response of y_t to the structural shocks can be uniquely decomposed into

$$\left(\frac{\partial y_t}{\partial \epsilon_t} \right)_{DSGE} = A_0(\theta) = \Sigma_{tr}^*(\theta) \Omega^*(\theta), \quad (40)$$

where $\Sigma_{tr}^*(\theta)$ is lower triangular and $\Omega^*(\theta)$ is orthonormal. To identify the VAR, we maintain the triangularization of its covariance matrix Σ_u and replace the rotation Ω in Equation (39) with the function $\Omega^*(\theta)$ that appears in (40).¹⁷

The implementation of this identification procedure is straightforward in our framework. Since we are able to generate draws from the joint posterior distribution of Φ , Σ_u , and θ , we can for each draw (i) use Φ to construct a MA representation of y_t in terms of the reduced-form shocks u_t , (ii) compute a Cholesky decomposition of Σ_u , and (iii) calculate $\Omega = \Omega^*(\theta)$ to obtain a MA representation in terms of the structural shocks ϵ_t .

A few remarks about the procedure are due. First, since the likelihood of the reduced form VAR is invariant with respect to Ω , the rotation matrix that is used to achieve identification is the same *a posteriori* as it is *a priori* for any fixed θ . However, as explained in Section 3.3.2, the distribution of θ is updated with the sample information. Hence, we learn from the data which rotation to choose, indirectly, via learning about the DSGE model parameters.

Second, the extent to which the posterior impulse responses are forced to look like the DSGE model's responses will depend on the tightness of the prior. The larger λ , the more similar the responses will be. If the researcher chooses the tightness of the prior endogenously based on the marginal data density (34) the data – and not

¹⁷Since the vector autoregressive representation of the DSGE model, as characterized by $\Phi^*(\theta)$ and $\Sigma_u^*(\theta)$, is only an approximation, the Cholesky decomposition of $\Sigma_u^*(\theta)$ is not exactly equal to $\Sigma_{tr}^*(\theta)$. However, in our experience the difference is for practical purposes negligible.

the researcher – will determine in which dimensions the posterior impulse responses will conform to the model’s responses, and in which dimensions they will not.

Figure 2 depicts the impulse response functions with respect to monetary policy shocks of cumulative real output growth, inflation, and the interest rate, normalized so that the initial impact of a monetary shock on the interest rate is 25 basis points. Each plot shows the VAR impulse-responses (dashed-and-dotted line), the corresponding 90 % error bands (dotted lines), and the DSGE model impulse responses (solid lines). The estimates are based on a sample of 80 observations ending in 2001:III. The impulse responses are computed for different values of the tightness parameter λ , namely $\lambda \in \{0.5, 1, 5\}$. As expected, the VAR impulse responses become closer to the model’s as the weight of the prior increases. Specifically, the distance between the posterior means of the VAR and the model’s impulse responses decreases. In addition, the bands for the VAR impulse responses narrow considerably.

It is interesting to observe that in some dimensions the VAR impulse responses conform to the model’s even for small values of the tightness parameter ($\lambda = 0.5$). The sign and the magnitude of the VAR impulse responses on impact agree with the model and are very precisely estimated. This suggests that Σ_{tr} and $\Sigma_{tr}^*(\theta)$ in Equations (39) and (40) are very similar. The response of inflation to a money shock is short-lived both in the DSGE model as well as in the VAR. In other dimensions there is less agreement: where the model predicts long-run money neutrality, the VAR impulse responses indicate that there is substantial uncertainty about the long-run effects of money shocks on output. While these findings are specific to this DSGE model, they seem to favor identification strategies based on impulse responses on impact (as in Faust (1998), Canova and DeNicoló (2002), and Uhlig (2001)) relative to strategies that rely on long-run neutrality.

Identification schemes based on zero-restrictions on the contemporaneous impact of the structural shocks often produce a price-puzzle in three- or four-variable VARs. While the price-puzzle can be avoided by including producer prices in addition to consumer prices, it can also be avoided by using our identification scheme that is

not based on zero-restrictions.

4.4 Regime Shifts

The analysis of welfare implications of different monetary policy rules has become an active area of research (see, for instance, the articles in Taylor (1999)). It is important for policy makers to have a set of tools that allows them to predict the effects of switching from one policy rule to another. Suppose a policy maker only has a DSGE model and an unrestricted VAR available. Moreover, suppose the forecast performance of the DSGE model is worse than that of the VAR. However, because of the Lucas critique, the policy maker does not fully trust the VAR to correctly predict the impact of a permanent change in the policy rule. Based on the DSGE-VAR, we propose an approach by which the policy maker uses the DSGE model to forecast the effect of the policy change, but corrects the DSGE model predictions. The correction is the one that brings the DSGE's forecasts under the existing policy regime on track. The tacit assumption underlying this procedure is that the correction is itself policy invariant.

We begin by decomposing the joint posterior of VAR and DSGE model parameters into

$$p(\Phi, \Sigma_u, \theta | Y) = p(\Phi, \Sigma_u | \theta, Y)p(\theta | Y). \quad (41)$$

To the extent that $p(\Phi, \Sigma_u | \theta, Y)$ assigns mass away from the restriction functions $\Phi^*(\theta)$ and $\Sigma_u^*(\theta)$ it can be interpreted as a correction of the VAR representation of the DSGE model given θ . This correction has been constructed from past observations to optimize forecasting performance.

Partition $\theta = [\theta'_{(s)}, \theta'_{(p)}]'$, where $\theta_{(p)}$ corresponds to the parameters that the central bank can affect with its conduct of monetary policy. Suppose that our beliefs about the post-intervention policy parameters can be characterized by the density $\tilde{p}(\theta_{(p)})$, which possibly concentrates all its mass on a single value of $\theta_{(p)}$. We combine $\tilde{p}(\theta_{(p)})$ with the information that we have about the non-policy parameters

based on the available data. This leads to a modified posterior of θ :

$$\tilde{p}(\theta|Y) = p(\theta_{(s)}|Y)\tilde{p}(\theta_{(p)}), \quad (42)$$

where $p(\theta_{(s)}|Y)$ is the marginal posterior of the non-policy parameters.

Our inference with respect to the effect of the regime shift will be drawn from

$$\tilde{p}_\lambda(\Phi, \Sigma_u|Y) = \int p_\lambda(\Phi, \Sigma_u|Y, \theta)\tilde{p}_\lambda(\theta|Y)d\theta. \quad (43)$$

We use the subscript λ to indicate that the conclusions depend on the weight given to the DSGE model. It is instructive to take another look at the posterior mean of the VAR parameters Φ conditional on the DSGE model parameters θ :

$$\tilde{\Phi}(\theta) = \left(\frac{\lambda}{1+\lambda} \Gamma_{xx}^*(\theta) + \frac{1}{1+\lambda} T^{-1} X' X \right)^{-1} \left(\frac{\lambda}{1+\lambda} \Gamma_{xy}^*(\theta) + \frac{1}{1+\lambda} T^{-1} X' Y \right).$$

The parameters θ are now drawn from the modified distribution $\tilde{p}(\theta|Y)$ that reflects the change in the distribution of the policy parameters due to the intervention. The sample moments $X'X$, $X'Y$, and $Y'Y$, on the other hand, are functions of pre-intervention observations only. They remain part of our analysis because they incorporate information about how to “correct” the DSGE model predictions to achieve a good forecast performance.

If $\lambda = 0$, the VAR posterior does not depend on θ at all: hence the researcher is ignoring the DSGE model (and the regime shift itself) in computing her forecasts. If $\lambda = \infty$, then the procedure is equivalent to analyzing the policy directly with the (VAR approximation of the) DSGE model. It is clear that the Lucas critique is fully observed only in the $\lambda = \infty$ extreme.

Most of the current literature on monetary policy rules focuses on the effect of these rules on the magnitude of economic fluctuations and the households’ utility over the business cycle. A popular measure of welfare besides agents’ utility is the volatility of the output gap and inflation. To illustrate our prediction approach we are considering the effect of a change in the response of the federal funds rate to deviations of inflation from its target rate (the parameter ψ_1 in the Taylor rule (14)) on the standard deviation of real output growth and inflation.

A widely shared belief, e.g., Clarida, Galí, and Gertler (2000), is that under the chairmanship of Paul Volcker and Alan Greenspan the U.S. central bank responded more aggressively to rising inflation than under their predecessors William Martin, Arthur Burns, and William Miller. Based on the empirical results in the Taylor-rule literature we compare two policies. Under the first policy scenario $\psi_1 = 1.1$, whereas under the second policy scenario $\psi_1 = 1.8$. The former can be loosely interpreted as a continuation of the inactive Martin, Burns, and Miller policy¹⁸, whereas the latter corresponds to a switch to a more active Volcker, Greenspan policy. To assess the two policies we generated draws from the modified posterior (42) and simulate trajectories of 80 observations conditional on the parameter draws. For each trajectory we discard the first eight quarters (hence we consider only the paths post 1982:III) and then compute the standard deviation of output growth, inflation, and the federal funds rate.

The results for various choices of λ are summarized in the density plots of Figure 3. The dashed densities corresponds to $\psi_1 = 1.1$ and the solid densities to $\psi_1 = 1.8$. The vertical lines in the plots show the standard deviation of the actual sample (post 1982:III) for the variables of interest. The 1982:III threshold is taken from Clarida, Galí, and Gertler (2000). The transition period from high inflation to low inflation between 1979 and 1982 implies that the actual standard deviation of inflation for the whole sample is high, and in our view does not reflect the “steady state” variability of inflation under the Volcker-Greenspan policy. Hence we choose to discard the first eight quarters.¹⁹

The DSGE model predicts that an increase in the Taylor rule parameter ψ_1 induces a lower equilibrium variability of inflation and therefore a lower variability

¹⁸Although some authors report estimates of $\psi_1 < 1$ we restrict ourselves to the determinacy region of the DSGE model.

¹⁹A possible explanation is that the linear model fails to capture the transition period from high inflation to low inflation. In the model agents change their expectations instantaneously when the policy change is announced whereas in reality there may be a learning process in which agents slowly realize that the policy change is permanent (regime shift) rather than temporary (deviation from policy rule).

of the federal funds rate. This is indeed what can be observed in Figure 3 for $\lambda > 0$. Whenever ψ_1 increases from 1.1 to 1.8, the forecasted variability of inflation and interest rate decreases. It is reassuring to observe that for inflation the distribution of the predicted standard deviation for $\psi_1 = 1.8$ (solid line) concentrates around the actual standard deviation. The predictions under $\psi_1 = 1.1$ (dashed line) tend to overestimate inflation variability. In terms of the interest rate, the predictions under $\psi_1 = 1.8$ tend to underestimate Fed Fund variability, while the predictions under $\psi_1 = 1.1$ tend to overestimate it (although the mode of the dashed distribution appears to be roughly on spot, the distribution is skewed to the right).

The predictions for the standard deviation of output do not change significantly as ψ_1 increases. The DSGE-VAR does not predict the reduction of the variability in real output growth that took place after 1979. This reduction may be the outcome of a change in the exogenous technology process rather than the effect of monetary policy. Note that the predicted effect of the policy change becomes larger as the weight of the prior increases. As more weight is given to the DSGE model, the differences in the predictions becomes sharper, as one would expect.

Although the experiment just described is not pure out-of-sample prediction, since the policy experiment $\psi_1 = 1.8$ was motivated by an analysis of the Volcker-Greenspan sample, it illustrates the potential of our approach. We view the procedure as a tool that lets the policy maker assess the effects of the policy change as a function of the confidence placed in the structural model measured by λ .

5 Conclusions

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, computationally efficient strategy, that is usable for policy analysis. Our approach involves the following steps: (i) Choose a DSGE model and a prior distribution for its parameters. (ii) Solve the DGSE model and map the prior distribution of its parameters into a prior distribution for the VAR parameters. While a log-linear approximation

of the DSGE model simplifies the computation of its VAR approximation given by $\Phi^*(\theta)$ and $\Sigma_u^*(\theta)$ considerably, it is not crucial to our approach. (iii) Obtain via Monte Carlo methods the joint posterior distribution of DSGE and VAR parameters, which can then be used to compute predictive densities.

We apply this procedure to a VAR in real output growth, inflation, and interest rates, and show that it is broadly successful in terms of forecasting performance. The DSGE-VAR clearly outperforms an unrestricted VAR and the vector autoregressive approximation of the DSGE model itself at all horizons. Its forecasting performance is comparable to, and in some dimensions better than a VAR with Minnesota prior.

Unlike the Minnesota prior, our DSGE model prior also offers help in terms of policy analysis. We construct a VAR identification scheme for the structural shocks based on a comparison of the contemporaneous VAR responses with the DSGE model responses. The identification of policy shocks enables an analysis of modest interventions. Our approach provides an attractive alternative to existing identification schemes because it closely ties VAR identification to a fully specified general equilibrium model. Beyond the specification of a DSGE model and the possibly data-driven selection of the prior weight, our identification procedure does not require the researcher to make any additional choices.

We also illustrate how a VAR with DSGE model prior can be used to predict the effects of a permanent change in the policy rule – a task that is generally considered infeasible for identified VARs. We use the approach to predict the impact of the change from the Martin-Burns-Miller regime to the Volcker-Greenspan regime on the volatility of the variables of interest. We find that at least in some dimensions the approach fares reasonably well in terms of predicting the effect of the change, although further research is needed to investigate this issue more deeply.

More research lies down the road. First, if the VAR is specified in terms of output and prices rather than output growth and inflation, then the calculation of expected sample moments of data generated from the DSGE model becomes more complicated due to the non-stationarity. Second, our method could be used to generate priors for state-space models that formally nest the DSGE model. This

extension is conceptually straightforward but challenging from a computational perspective. Third, it is worthwhile to apply the method to larger scale DSGE models, e.g., the Smets-Wouters (2003) model, and to make comparisons among priors that are derived from different models, such as a New Keynesian model versus a flexible price cash-in-advance model.

As envisioned in Diebold (1998), the combination of DSGE models and vector autoregressions shows promise for macroeconomic forecasting and policy analysis.

References

- Altug, Sumru (1989): “Time-To-Build and Aggregate Fluctuations: Some New Evidence.” *International Economic Review*, **30**(4), 889-921.
- Canova, Fabio and Gianni De Nicoló (2002): “Monetary Disturbances Matter for Business Cycle Fluctuations in the G-7.” *Journal of Monetary Economics*, **49**(6), 1131-1159.
- Chamberlain, Gary (1984): “Panel Data.” In: Z. Griliches and M. Intriligator (eds.), *Handbook of Econometrics*, Vol. 3, New York, 1465-1514.
- Christiano, Lawrence, Martin Eichenbaum, and Charles Evans (1999): “Monetary Policy Shocks: What Have We Learned and to What End?” In: John Taylor and Michael Woodford (eds.), *Handbook of Macroeconomics*, Elsevier Science, New York.
- Clarida, Richard, Jordi Galí, and Mark Gertler (2000): “Monetary Policy Rules and Macroeconomic Stability: Evidence and some Theory.” *Quarterly Journal of Economics*, **115**, 147-180.
- Diebold, Francis X. (1998): “The Past, Present, and Future of Macroeconomic Forecasting.” *Journal of Economic Perspectives*, **12**, 175-192.
- Diebold, Francis X., and Roberto S. Mariano (1995): “Comparing Predictive Accuracy.” *Journal of Business & Economic Statistics*, **13**, 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.
- Faust, Jon (1998): "The Robustness of Identified VAR Conclusions about Money." *Carnegie Rochester Conference Series*, **49**, 207-244.
- Geweke, John (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communications." *Econometrics Reviews*, **18**, 1-127.
- 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.
- Ireland, Peter N. (2003): "A Method for Taking Models to the Data." *Journal of Economic Dynamics and Control*, forthcoming.
- King, Robert G. (2000): "The New IS-LM Model: Language, Logic, and Limits." Federal Reserve Bank of Richmond *Economic Quarterly*, **86**, 45-103.
- King, Robert G., Charles I. Plosser and Sergio T. Rebelo (1988): "Production, Growth and Business Cycles." *Journal of Monetary Economics*, **21**, 195-232.
- King, Robert G. and Alexander L. Wolman (1999): "What Should the Monetary Authority Do If Prices Are Sticky?" In: John B. Taylor (ed.), *Monetary Policy Rules*. University of Chicago Press.
- Leeper, Eric M., and Tao Zha (2003): "Modest Policy Interventions." *Journal of Monetary Economics*, forthcoming.
- Lehmann, Erich (1983): "*Theory of Point Estimation.*" John Wiley & Sons, New York.

- Lubik, Thomas and Frank Schorfheide (2002): "Testing for Indeterminacy: An Application to U.S. Monetary Policy." *Manuscript*, Department of Economics, University of Pennsylvania.
- Litterman, Robert B. (1986): "Forecasting with Bayesian Vector Autoregressions – Five Years of Experience." *Journal of Business & Economic Statistics*, **4**, 25-37.
- Lucas, Robert E. (1976): "Econometric Policy Evaluation: A Critique." In: K. Brunner and A. Meltzer (eds.), *The Phillips Curve and the Labor Market*, Carnegie-Rochester Conference Series, **1**, Amsterdam: North-Holland.
- Moon, Hyungsik Roger and Frank Schorfheide (2002): "Minimum Distance Estimation of Non-stationary Time Series Models." *Econometric Theory*, **18**, 1385-1407.
- Robertson, John C., Ellis W. Tallman, and Charles H. Whiteman (2002): "Forecasting Using Relative Entropy." *Federal Reserve Bank of Atlanta Working Paper*, **2002-22**.
- Sargent, Thomas J. (1989): "Two Models of Measurements and the Investment Accelerator." *Journal of Political Economy*, **97**(2), 251-287.
- Schorfheide, Frank (2000): "Loss Function-Based Evaluation of DSGE Models." *Journal of Applied Econometrics*, **15**, 645-670.
- Sims, Christopher A. (1980): "Macroeconomics and Reality." *Econometrica*, **48**, 1-48.
- (2002): "Solving Rational Expectations Models." *Computational Economics*, **20**(1-2), 1-20.
- Sims, Christopher A. and Tao Zha (1998): "Bayesian methods for dynamic multivariate models." *International Economic Review*, **39**, 949-968.

- Smets, Frank and Rafael Wouters (2003): “An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area.” *Journal of the European Economic Association*, forthcoming.
- Smith, Anthony A. Jr. (1993): “Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions.” *Journal of Applied Econometrics*, **8**, S63-S84.
- Taylor, John B. (1999): “*Monetary Policy Rules*.” University of Chicago Press, Chicago.
- Theil, Henry and Arthur S. Goldberger (1961): “On Pure and Mixed Estimation in Economics.” *International Economic Review*, **2**, 65-78.
- Todd, Richard M. (1984): “Improving Economic Forecasting with Bayesian Vector Autoregression.” *Federal Reserve Bank of Minneapolis Quarterly Review*, 18-29.
- Uhlig, Harald (2001): “What are the Effects of Monetary Policy? Evidence from an Agnostic Identification Procedure.” *Manuscript*, Humboldt University.
- Woodford, Michael (2003): “*Interest and Prices: Foundations of a Theory of Monetary Policy*.” Forthcoming, Princeton University Press, Princeton.
- Zellner, Arnold (1971): “*Introduction to Bayesian Inference in Econometrics*.” John Wiley & Sons, New York.

A Derivations and Proofs

The normalization factor $c(\theta)$ for the conditional prior density of the VAR parameters $p(\Phi, \Sigma_u | \theta)$ in Equation (21) is given by

$$\begin{aligned} c(\theta) = & (2\pi)^{\frac{nk}{2}} |\lambda T \Gamma_{xx}^*(\theta)|^{-\frac{n}{2}} |\lambda T \Sigma_u^*(\theta)|^{-\frac{\lambda T - k}{2}} \\ & 2^{\frac{n(\lambda T - k)}{2}} \pi^{\frac{n(n-1)}{4}} \prod_{i=1}^n \Gamma[(\lambda T - k + 1 - i)/2], \end{aligned} \quad (\text{A1})$$

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

The marginal likelihood function of θ in Equation (32) is given by

$$\begin{aligned} p(Y|\theta) &= p(Y|\Phi, \Sigma)p(\Phi, \Sigma|\theta)/p(\Phi, \Sigma|Y) \\ &= \frac{|\lambda T\Gamma_{xx}^*(\theta) + X'X|^{-\frac{n}{2}} |(\lambda+1)T\tilde{\Sigma}_u(\theta)|^{-\frac{(\lambda+1)T-k}{2}}}{|\lambda T\Gamma_{xx}^*(\theta)|^{-\frac{n}{2}} |\lambda T\Sigma_u^*(\theta)|^{-\frac{\lambda T-k}{2}}} \\ &\quad \times \frac{(2\pi)^{-nT/2} 2^{\frac{n((\lambda+1)T-k)}{2}} \prod_{i=1}^n \Gamma[((\lambda+1)T-k+1-i)/2]}{2^{\frac{n(\lambda T-k)}{2}} \prod_{i=1}^n \Gamma[(\lambda T-k+1-i)/2]}. \end{aligned} \quad (\text{A2})$$

The third equality can be obtained from the normalization constants of the Inverted Wishart – Normal distributions.

Proof of Proposition 1. Define the sample moments $\hat{\Gamma}_{xx} = X'X/T$, $\hat{\Gamma}_{xy} = X'Y/T$, and $\hat{\Gamma}_{yy} = Y'Y/T$. Let $\phi = 1/\lambda$ and $\tilde{\theta}$ the mode of the marginal log-likelihood function given in Equation (A2). Consider the log-likelihood ratio

$$\begin{aligned} \ln \frac{p(Y|\theta)}{p(Y|\tilde{\theta})} &= -\frac{T}{2} \ln |\Sigma_u^*(\theta)| - \frac{n}{2} \ln |\mathcal{I} + \phi\Gamma_{xx}^{*-1}(\theta)\hat{\Gamma}_{xx}| \\ &\quad - \frac{(1/\phi+1)T-k}{2} \ln \left| \frac{1/\phi+1}{1/\phi} \Sigma_u^{*-1}(\theta)\tilde{\Sigma}_u(\theta) \right| \\ &\quad + \frac{T}{2} \ln |\Sigma_u^*(\tilde{\theta})| + \frac{n}{2} \ln |\mathcal{I} + \phi\Gamma_{xx}^{*-1}(\tilde{\theta})\hat{\Gamma}_{xx}| \\ &\quad + \frac{(1/\phi+1)T-k}{2} \ln \left| \frac{1/\phi+1}{1/\phi} \Sigma_u^{*-1}(\tilde{\theta})\tilde{\Sigma}_u(\tilde{\theta}) \right|. \end{aligned} \quad (\text{A3})$$

We derive an approximation of the log-likelihood ratio that is valid as $\phi \rightarrow 0$. A first-order Taylor approximation of the second term around $\phi = 0$ yields

$$\ln |\mathcal{I} + \phi\Gamma_{xx}^{*-1}\hat{\Gamma}_{xx}| = \ln |\mathcal{I}| + \phi \text{tr}[\Gamma_{xx}^{*-1}\hat{\Gamma}_{xx}] + O(\phi^2). \quad (\text{A4})$$

Notice that

$$\begin{aligned} \frac{1/\phi+1}{1/\phi} \Sigma_u^{*-1} \tilde{\Sigma}_u &= \left[\Gamma_{yy}^* - \Gamma_{yx}^* \Gamma_{xx}^{*-1} \Gamma_{xy}^* \right]^{-1} \\ &\quad \times \left[\Gamma_{yy}^* + \phi \hat{\Gamma}_{yy} - (\Gamma_{yx}^* + \phi \hat{\Gamma}_{yx})(\Gamma_{xx}^* + \phi \hat{\Gamma}_{xx})^{-1} (\Gamma_{xy}^* + \phi \hat{\Gamma}_{xy}) \right]. \end{aligned} \quad (\text{A5})$$

The log-determinant of this term has the following first-order Taylor expansion around $\phi = 0$:

$$\begin{aligned} \ln \left| \frac{1/\phi+1}{1/\phi} \Sigma_u^{*-1} \tilde{\Sigma}_u \right| \\ = \ln |\mathcal{I}| + \phi \text{tr} \left[\Sigma_u^{*-1} (\hat{\Gamma}_{yy} - \hat{\Gamma}_{yx} \Phi^* - \Phi^{*\prime} \hat{\Gamma}_{xy} + \Phi^{*\prime} \hat{\Gamma}_{xx} \Phi^*) \right] + O(\phi^2). \end{aligned} \quad (\text{A6})$$

Combining these results yields

$$\begin{aligned}
\ln p(Y|\theta) &= -\frac{T}{2} \ln |\Sigma_u^*(\theta)| + \frac{T}{2} \ln |\Sigma_u^*(\tilde{\theta})| \\
&\quad - \frac{T}{2} \text{tr} \left[\Sigma_u^{*-1}(\theta) (\hat{\Gamma}_{yy} - \hat{\Gamma}_{yx}\Phi^*(\theta) - \Phi^{*\prime}(\theta)\hat{\Gamma}_{xy} + \Phi^{*\prime}(\theta)\hat{\Gamma}_{xx}\Phi^*(\theta)) \right] \\
&\quad + \frac{T}{2} \text{tr} \left[\Sigma_u^{*-1}(\tilde{\theta}) (\hat{\Gamma}_{yy} - \hat{\Gamma}_{yx}\Phi^*(\tilde{\theta}) - \Phi^{*\prime}(\tilde{\theta})\hat{\Gamma}_{xy} + \Phi^{*\prime}(\tilde{\theta})\hat{\Gamma}_{xx}\Phi^*(\tilde{\theta})) \right] + O(\phi) \\
&= \ln \frac{p^*(Y|\theta)}{p^*(Y|\tilde{\theta})} + O(\phi).
\end{aligned} \tag{A7}$$

Thus, as $\phi \rightarrow 0$ the log-likelihood ratio converges to the log-likelihood ratio of the quasi-likelihood functions. The convergence is uniform on compact subsets of Θ for which $\Sigma_u^*(\theta)$ and $\Gamma_{xx}^*(\theta)$ are non-singular.

Proof of Proposition 2. We rewrite the marginal log-likelihood ratio given in Equation (A3) in terms of λ :

$$\begin{aligned}
\ln \frac{p(Y|\theta)}{p(Y|\tilde{\theta})} &= -\frac{T}{2} \ln |\Sigma_u^*(\theta)| - \frac{n}{2} \ln |\lambda \mathcal{I} + \Gamma_{xx}^{*-1}(\theta) \hat{\Gamma}_{xx}| \\
&\quad - \frac{(\lambda+1)T-k}{2} \ln |(\lambda+1)\Sigma_u^{*-1}(\theta) \tilde{\Sigma}_u(\theta)| \\
&\quad + \frac{T}{2} \ln |\Sigma_u^*(\tilde{\theta})| + \frac{n}{2} \ln |\lambda \mathcal{I} + \Gamma_{xx}^{*-1}(\tilde{\theta}) \hat{\Gamma}_{xx}| \\
&\quad + \frac{(\lambda+1)T-k}{2} \ln |(\lambda+1)\Sigma_u^{*-1}(\tilde{\theta}) \tilde{\Sigma}_u(\tilde{\theta})|.
\end{aligned} \tag{A8}$$

A Taylor-series expansion of the second term around $\lambda = 0$ yields

$$\ln |\lambda \mathcal{I} + \Gamma_{xx}^{*-1} \hat{\Gamma}_{xx}| = \ln |\Gamma_{xx}^{*-1} \hat{\Gamma}_{xx}| + \lambda \text{tr}[\hat{\Gamma}_{xx}^{-1} \Gamma_{xx}^*] + O(\lambda^2). \tag{A9}$$

Notice that

$$\begin{aligned}
(\lambda+1)\Sigma_u^{*-1} \tilde{\Sigma}_u &= \left[\Gamma_{yy}^* - \Gamma_{yx}^* \Gamma_{xx}^{*-1} \Gamma_{xy}^* \right]^{-1} \\
&\quad \times \left[\lambda \Gamma_{yy}^* + \hat{\Gamma}_{yy} - (\lambda \Gamma_{yx}^* + \hat{\Gamma}_{yx})(\lambda \Gamma_{xx}^* + \hat{\Gamma}_{xx})^{-1} (\lambda \Gamma_{xy}^* + \hat{\Gamma}_{xy}) \right].
\end{aligned} \tag{A10}$$

The log-determinant of this term has the following first-order Taylor expansion around $\lambda = 0$:

$$\begin{aligned}
&\ln |(\lambda+1)\Sigma_u^{*-1} \tilde{\Sigma}_u| \\
&= \ln |\Sigma_u^{*-1} \hat{\Sigma}_{u,mle}| + \lambda \text{tr} \left[\hat{\Sigma}_{u,mle}^{-1} (\Gamma_{yy}^* - \Gamma_{yx}^* \hat{\Phi}_{mle} - \hat{\Phi}'_{mle} \Gamma_{xy}^* + \hat{\Phi}'_{mle} \Gamma_{xx}^* \hat{\Phi}_{mle}) \right] + O(\lambda^2) \\
&= \ln |\Sigma_u^{*-1} \hat{\Sigma}_{u,mle}| + \lambda \text{tr} \left[\hat{\Sigma}_{u,mle}^{-1} \Sigma_u^* \right] + \lambda \text{tr} \left[\hat{\Sigma}_{u,mle}^{-1} (\Phi^* - \hat{\Phi}_{mle})' \Gamma_{xx}^* (\Phi^* - \hat{\Phi}_{mle}) \right] + O(\lambda^2).
\end{aligned} \tag{A11}$$

Combining the three terms leads to the following approximation of the log-likelihood ratio

$$\begin{aligned}
\frac{p(Y|\theta)}{p(Y|\tilde{\theta})} &= -\frac{\lambda T}{2} \ln |\Sigma_u^{*-1}(\theta)| + \frac{\lambda T}{2} \ln |\Sigma_u^{*-1}(\tilde{\theta})| \\
&\quad - \frac{\lambda T}{2} \text{tr}[\hat{\Sigma}_{u,mle}^{-1} \Sigma_u^*(\theta)] + \frac{\lambda T}{2} \text{tr}[\hat{\Sigma}_{u,mle}^{-1} \Sigma_u^*(\tilde{\theta})] \\
&\quad - \frac{\lambda T}{2} \text{tr}[\hat{\Sigma}_{u,mle}^{-1} (\Phi^*(\theta) - \hat{\Phi}_{mle})' \Gamma_{xx}^*(\theta) (\Phi^*(\theta) - \hat{\Phi}_{mle})] \\
&\quad + \frac{\lambda T}{2} \text{tr}[\hat{\Sigma}_{u,mle}^{-1} (\Phi^*(\tilde{\theta}) - \hat{\Phi}_{mle})' \Gamma_{xx}^*(\tilde{\theta}) (\Phi^*(\tilde{\theta}) - \hat{\Phi}_{mle})] + O_p(\max[\lambda^2 T, 1]) \\
&= \ln \frac{q(Y|\theta)}{q(Y|\tilde{\theta})} + O_p(\max[\lambda^2 T, 1]).
\end{aligned} \tag{A12}$$

B Practical Implementation

We assume that the parameter space of λ is finite $\Lambda = \{l_1, \dots, l_q\}$. In order to select λ , and to generate draws from the joint posterior distribution of DSGE model parameters and VAR parameters we use the following scheme:

1. For each $\lambda \in \Lambda$ use the Metropolis algorithm described in Schorfheide (2000) to generate draws from $p_\lambda(\theta|Y) \propto p_\lambda(Y|\theta)p(\theta)$. The steps needed to evaluate $p_\lambda(Y|\theta)$ based on Equation (A2) are as follows. For each θ :
 - (i) Solve the DSGE model given by Equations (7), (10), and (12) to (14), for instance, with the algorithm described in Sims (2002). This leads to a transition equation of the form

$$s_t = T(\theta)s_{t-1} + R(\theta)\epsilon_t. \tag{B1}$$

The measurement equation (15) can be written in stacked form as:

$$y_t = Z(\theta)s_t + D(\theta) + \nu_t. \tag{B2}$$

(In our implementation we choose s_t such that $\nu_t = 0$). Define the variance-covariance matrices of the shocks as:

$$\mathbb{E}[\nu_t \nu_t'] = \Sigma_{\nu\nu}(\theta), \quad \mathbb{E}[\epsilon_t \epsilon_t'] = \Sigma_{\epsilon\epsilon}(\theta), \quad \mathbb{E}[\epsilon_t \nu_t'] = \Sigma_{\epsilon\nu}(\theta).$$

- (ii) Compute the population moments $\Gamma_{yy}^*(\theta)$, $\Gamma_{yx}^*(\theta)$, and $\Gamma_{xx}^*(\theta)$ from the state-space representation (B1, B2). Notice that

$$\begin{aligned}\mathbb{E}[y_t y'_t] &= Z \Omega_{ss} Z' + Z R \Sigma_{\epsilon\nu} + (Z R \Sigma_{\epsilon\nu})' + \Sigma_{\nu\nu} + D D' \\ \mathbb{E}[y_t y'_{t-h}] &= Z T^h (\Omega_{ss} Z' + R \Sigma_{\epsilon\nu}) + D D'\end{aligned}\tag{B3}$$

where $\Omega_{ss} = \mathbb{E}[s_t s'_t]$ and can be obtained by solving the following Lyapunov equation $\Omega_{ss} = T \Omega_{ss} T' + R \Sigma_{\epsilon\epsilon} R'$.

2. Based on these draws apply Geweke's (1999) modified harmonic mean estimator to obtain numerical approximations of the data densities $p_\lambda(Y)$.
3. Find the pre-sample size $\hat{\lambda}$ that has the highest data density.
4. Select the draws of $\{\theta_{(s)}\}$ that correspond to $\hat{\lambda}$ and use standard methods to generate draws from $p(\Phi, \Sigma_u | Y, \theta_{(s)})$ (see Equations 30 and 31) for each $\theta_{(s)}$.

Notice that this scheme can also be used to select among competing DSGE models. Moreover, the whole procedure can be easily generalized to the case in which we have a prior distribution over the hyperparameter λ .

Table 1: PRIOR DISTRIBUTIONS FOR DSGE MODEL PARAMETERS

Name	Range	Density	Mean	S.D.
$\ln \gamma$	\mathbb{R}	Normal	0.500	0.250
$\ln \pi^*$	\mathbb{R}	Normal	1.000	0.500
$\ln r^*$	\mathbb{R}^+	Gamma	0.500	0.250
κ	\mathbb{R}^+	Gamma	0.300	0.150
τ	\mathbb{R}^+	Gamma	2.000	0.500
ψ_1	\mathbb{R}^+	Gamma	1.500	0.250
ψ_2	\mathbb{R}^+	Gamma	0.125	0.100
ρ_R	[0,1)	Beta	0.500	0.200
ρ_g	[0,1)	Beta	0.800	0.100
ρ_z	[0,1)	Beta	0.300	0.100
σ_R	\mathbb{R}^+	Inv. Gamma	0.251	0.139
σ_g	\mathbb{R}^+	Inv. Gamma	0.630	0.323
σ_z	\mathbb{R}^+	Inv. Gamma	0.875	0.430

Notes: The model parameters $\ln \gamma$, $\ln \pi^*$, $\ln r^*$, σ_R , σ_g , and σ_z are scaled by 100 to convert them into percentages. The Inverse Gamma priors are of the form $p(\sigma|\nu, s) \propto \sigma^{-\nu-1} e^{-\nu s^2/2\sigma^2}$, where $\nu = 4$ and s equals 0.2, 0.5, and 0.7, respectively. Approximately 1.5 % of the prior mass lies in the indeterminacy region of the parameter space. The prior is truncated in order to restrict it to the determinacy region of the DSGE model.

Table 2: POSTERIOR OF DSGE MODEL PARAMETERS: 1959:III - 1979:II

Name	Prior		Posterior, $\lambda = 1$		Posterior, $\lambda = 10$	
	CI(low)	CI(high)	CI(low)	CI(high)	CI(low)	CI(high)
$\ln \gamma$	0.101	0.922	0.473	1.021	0.616	1.045
$\ln \pi^*$	0.219	1.863	0.433	1.613	0.553	1.678
$\ln r^*$	0.132	0.880	0.113	0.463	0.126	0.384
κ	0.063	0.513	0.101	0.516	0.081	0.416
τ	1.197	2.788	1.336	2.816	1.684	3.225
ψ_1	1.121	1.910	1.011	1.559	1.009	1.512
ψ_2	0.001	0.260	0.120	0.497	0.150	0.545
ρ_R	0.157	0.812	0.530	0.756	0.550	0.747

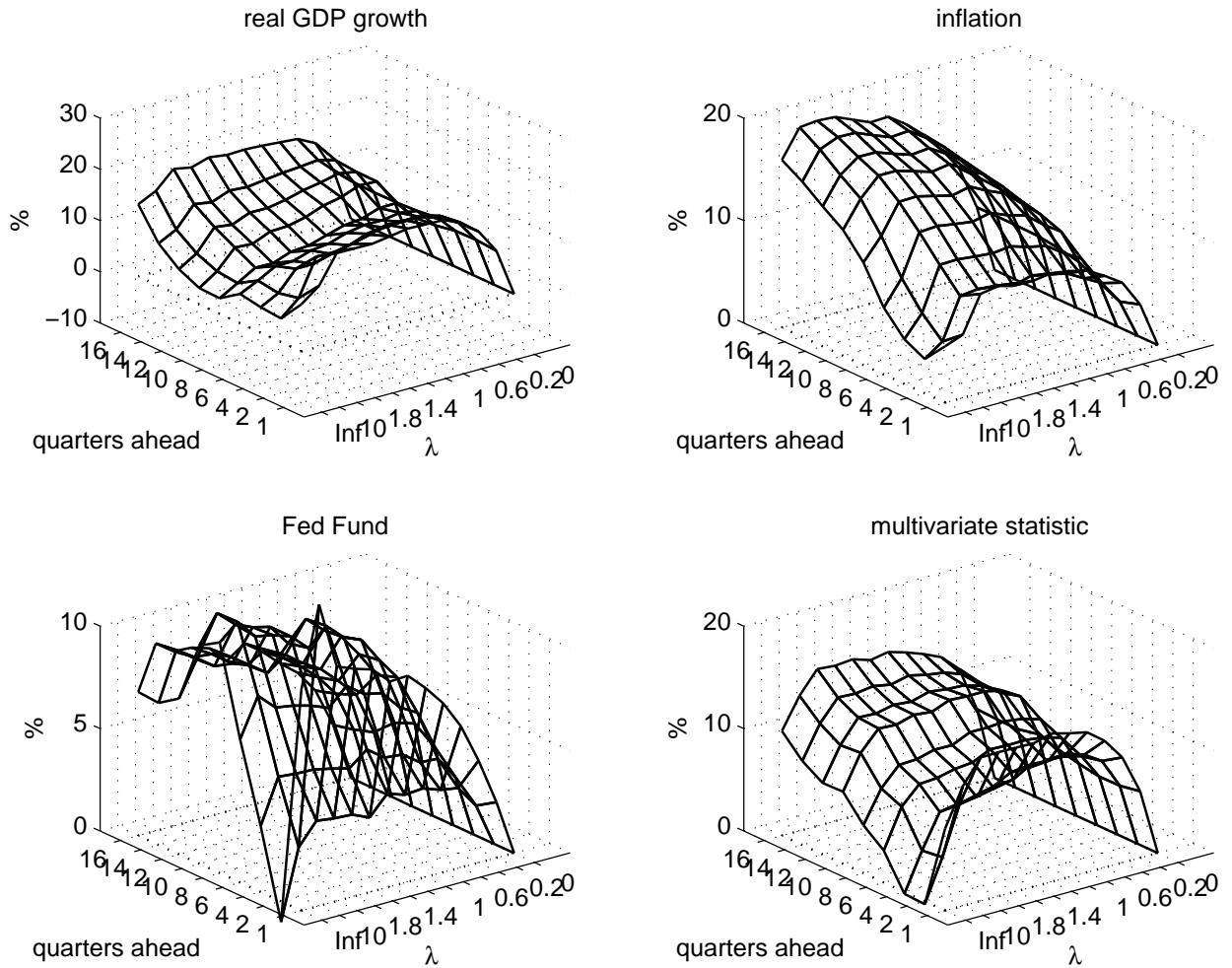
Notes: We report 90 % confidence intervals based on the output of the Metropolis-Hastings Algorithm. The model parameters $\ln \gamma$, $\ln \pi^*$, and $\ln r^*$ are scaled by 100 to converted them into percentages.

Table 3: PERCENTAGE GAIN (LOSS) IN RMSEs: DSGE PRIOR VERSUS UNRESTRICTED VAR AND MINNESOTA PRIOR

Horizon	rGDP Growth		Inflation		Fed Funds		Multivariate	
	V-unr	V-Minn	V-unr	V-Minn	V-unr	V-Minn	V-unr	V-Minn
1	17.335	1.072	8.389	1.653	7.250	-7.593	12.842	0.943
2	16.977	6.965	7.247	1.339	5.024	-4.895	10.993	2.884
4	15.057	5.803	8.761	4.767	5.008	-1.878	9.630	3.959
6	14.116	3.452	10.460	7.240	6.648	-0.713	10.388	4.290
8	12.387	4.230	11.481	7.794	8.420	-0.204	11.023	5.187
10	14.418	7.986	12.261	8.351	8.242	-0.639	12.864	6.463
12	15.078	12.512	12.626	9.011	6.404	0.726	12.419	7.537
14	16.236	17.233	12.995	9.634	6.059	1.146	12.611	8.481
16	19.122	21.575	13.238	10.116	5.823	2.389	13.428	9.512

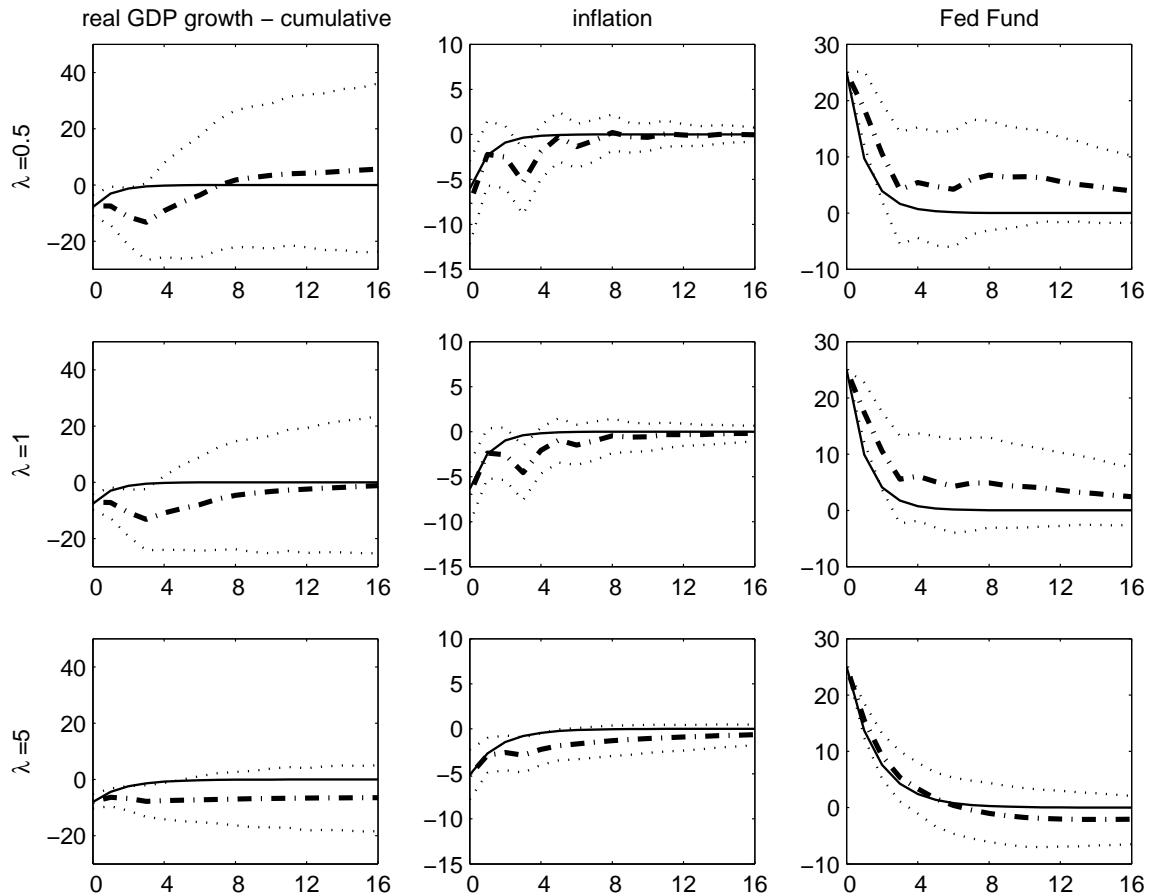
Notes: The rolling sample is 1975:III to 1997:III (90 periods). At each date in the sample, 80 observations are used in order to estimate the VAR. The forecasts are computed based on the values $\hat{\lambda}$ and \hat{i} that have the highest posterior probability based on the estimation sample.

Figure 1: FORECASTING PERFORMANCE AS A FUNCTION OF THE WEIGHT OF THE PRIOR



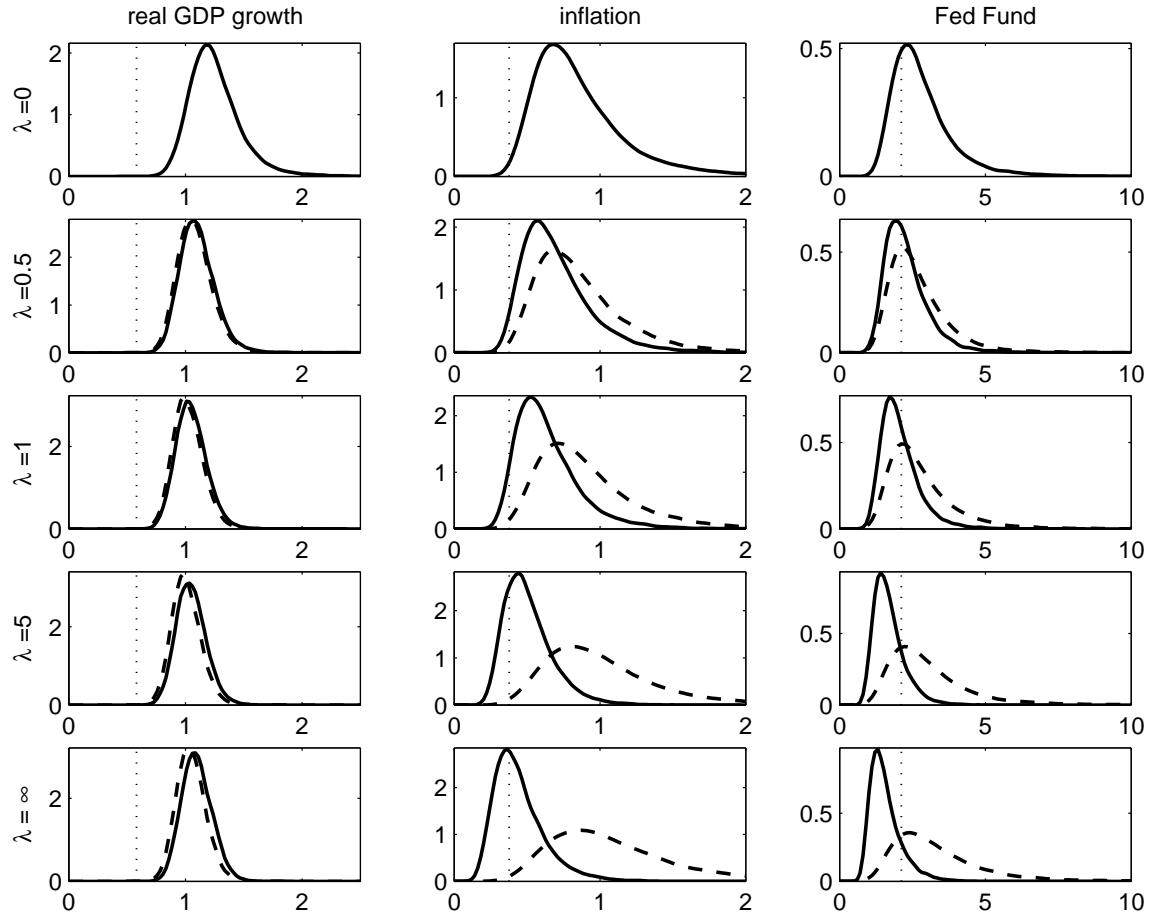
Notes: The plot shows the percentage gain (loss) in RMSEs relative to an unrestricted VAR. The rolling sample is 1975:III to 1997:III (90 periods). At each date in the sample, 80 observations are used in order to estimate the VAR.

Figure 2: IDENTIFIED IMPULSE RESPONSE FUNCTIONS



Notes: The dashed-dotted lines represent the posterior means of the VAR impulse response functions. The dotted lines are 90% confidence bands. The solid lines represent the mean impulse responses from the DSGE model. The impulse responses are based on the sample 1981:IV to 2001:III.

Figure 3: EFFECTS OF A POLICY REGIME SHIFT



Notes: The dotted horizontal lines correspond to the sample standard deviation of the actual data from 1982:IV to 1999:II. The dashed and the solid lines are posterior predictive distributions of sample standard deviations for the same time period, obtained using data up to 1979:II. The dashed line corresponds to $\psi_1 = 1.1$, the solid line corresponds to $\psi_1 = 1.8$.