Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://www.econ.upenn.edu/pier

PIER Working Paper 02-024

"Priors from General Equilibrium Models for VARs"

by

Marco Del Negro and Frank Schorfheide

Priors from General Equilibrium Models for VARs*

Marco Del Negro
Research Department
Federal Reserve Bank of Atlanta†

Frank Schorfheide
Department of Economics
University of Pennsylvania‡

July 2002

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 Sean Campbell, Frank Diebold, John Geweke, Thomas Lubik, Chris Sims, Dan Waggoner, Charles Whiteman, Tao Zha, seminar participants at Duke University, the European Central Bank, the Federal Reserve Bank of Atlanta, the Federal Reserve System Meetings, the 2002 Midwest Macro Conference, the 2002 CIRANO Forecasting Conference, and the University of Pennsylvania for helpful comments and discussions. 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 monetary DSGE model as a prior for a VAR, 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 Ph.D. course, 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 top 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 offering 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 (2002), this paper proposes an approach that combines a stylized general equilibrium model with a vector autoregression 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. We will specify a hierarchical prior starting out with a distribution for the structural DSGE model parameters. 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

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

VAR to these data. In practice we replace the sample moments of the simulated data by population moments computed from the DSGE model solution. 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 Ingram and Whiteman (1994) and DeJong, Ingram, and Whiteman (1993), who are the first to use priors from DSGE models for VARs. Ingram and Whiteman showed that prior information from the bare-bone 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 a DSGE model based prior distribution for the VAR impulse responses, which can be updated through the sample information.

[2] Ireland (1999) pursues a similar goal with a very different approach. He augments the linearized solution of a DSGE model with unobservable errors that have a VAR representation. We do not directly compare the forecasting accuracy of the two approaches. Since his model hinges on unobservables, which may or may not contain policy shocks, it is less suitable than the approach pursued here for policy experiments. In addition our approach has the advantage that one can control the relative weight of DSGE model and VAR in the hybrid model.

The second policy experiment is more ambitious. We want to forecast the effects of a change in the policy rule. Loosely speaking, our approach can be seen as a weighted average between (i) using the DSGE model only to forecast the effects of the policy change, and (ii) using the VAR only to make forecasts, thereby ignoring the effect of the policy change on the economic dynamics. The choice of the weight is tied to the confidence that we place on the structural model conditional on the observed data. As an application for 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. The results suggest that the approach is promising, and superior the extremes (i) and (ii).

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 variant of what is often referred to as New Keynesian IS-LM model. To make this paper self-contained, we briefly review the model specification, which is adopted from Lubik and Schorfheide (2002). Related descriptions and detailed derivations can be found, among others, in Galí and Gertler (1999), King (2000), King and Wolman (1999), and Woodford (2000).

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 consumption c and real balances M/P, and disutility from working:

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

where h denotes hours worked, β is the discount factor, τ is the risk aversion, 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. Consequently, the household maximizes (1) subject to its budget constraint:

$$c_t + \frac{B_t}{P_t} + \frac{M_t}{P_t} + \frac{T_t}{P_t} = w_t h_t + \frac{M_{t-1}}{P_t} + R_{t-1}\frac{B_{t-1}}{P_t} + D_t. \tag{2}$$

The usual transversality condition on asset accumulation rules out Ponzi-schemes. Initial conditions are given by B_0.

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

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

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:

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

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:

$$x_t(j) = z_t h_t(j). \tag{5}$$

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

$$\Delta \ln z_t = (1 - \rho_z) \ln \gamma + \rho_z \Delta \ln z_{t-1} + \epsilon_{z,t}, \tag{6}$$

where Δ denotes the temporal difference operator 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. Since our simple DSGE model lacks an internal propagation mechanism that can generate serially correlated output growth rates we assume that $\Delta \ln z_t$ follows a stationary AR(1) process.

Firm j chooses its labor input $h_t(j)$ and price $P_t(j)$ I to maximize

$$\mathbb{E}_t \left[\sum_{s=t}^{\infty} q_s D_s(j) \right] \tag{7}$$

subject to (5) and (6), where

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

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 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 output and inflation from their respective target levels:

$$R_t = f_t(\pi_t, x_t, R_{t-1}, \epsilon_{R,t}). \tag{8}$$

The central bank supplies the money demanded by the household. The monetary policy shock $\epsilon_{R,t}$ can be interpreted as an unanticipated deviation from the policy rule.

To complete the specification of the model it is assumed that 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):

$$\frac{T_t}{P_t} - \frac{M_t - M_{t-1}}{P_t} + \frac{B_t - R_{t-1}B_{t-1}}{P_t} = \zeta_t x_t. \tag{9}$$

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. For simplicity we assume that the government consumes a fraction ζ_t of each individual good j. We define $g_t = 1/(1-\zeta_t)$ and assume that g_t follows a stationary AR(1) process

$$\ln g_t = (1-\rho_g)\ln g^* + \rho_g \ln g_{t-1} + \epsilon_{g,t}, \tag{10}$$

where $\epsilon_{g,t}$ can be broadly interpreted as government spending shock.

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 z_t. The model has a deterministic steady state in terms of the detrended variables. To approximate the equilibrium dynamics, the model is log-linearized and the resulting linear rational expectations system is solved with the algorithm described in Sims (2000).

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

$$\begin{aligned}
\widetilde{x}_t &= I\!E_t[\widetilde{x}_{t+1}] - \tau^{-1}(\widetilde{R}_t - I\!E_t[\widetilde{\pi}_{t+1}]) + (1-\rho_g)\widetilde{g}_t + \rho_z \Delta\widetilde{z}_t, && (11)\\
\widetilde{\pi}_t &= \beta\gamma^{1-\tau} I\!E_t[\widetilde{\pi}_{t+1}] + \kappa[\widetilde{x}_t - \widetilde{g}_t], && (12)\\
\widetilde{R}_t &= \rho_R \widetilde{R}_{t-1} + (1-\rho_R)\left(\psi_1\widetilde{\pi}_t + \psi_2\widetilde{x}_t\right) + \epsilon_{R,t}, && (13)
\end{aligned}$$

where 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 (11), often referred to as the New Keynesian IS-curve, is an intertemporal Euler-equation, while (12) is derived from the firms' optimal price-setting problem and governs inflation dynamics around the steady state π. This relation can be interpreted as an (expectational) Phillips-curve with slope κ. Equation (13) is the log-linearized monetary policy rule where $0 \geq \rho_R < 1$ is the smoothing coefficient and ψ_1, ψ_2 are the elasticities of the target interest rate with respect to the deviation of inflation and output from their targets.[3]

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

$$\begin{aligned} \Delta \ln x_t &= \ln \gamma + \Delta \ln \widetilde{x}_t + \Delta \widetilde{z}_t \\ \Delta \ln P_t &= \ln \pi^* + \widetilde{\pi}_t \\ \ln R_t^a &= 4[(\ln r^* + \ln \pi^*) + \widetilde{R}_t], \end{aligned} \tag{14}$$

where the steady-state real interest rate $r^* = 1/\beta$. 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 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 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

$$\theta = [\ln \gamma, \ln \pi^*, \ln r^*, \kappa, \tau, \psi_1, \psi_2, \rho_R, \rho_g, \rho_z, \sigma_R, \sigma_g, \sigma_z]'. \tag{15}$$

3 A VAR Prior 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 $y_t = [\Delta \ln x_t, \Delta \ln P_t, \ln R_t^a]$. The VAR model is of

[3] In 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.

the form

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \dots \Phi_p y_{t-p} + u_t, \quad u_t \sim \mathcal{N}(0, \Sigma_u), \tag{16}$$

where u_t is a vector of reduced-form disturbances. 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

$$Y = X\Phi + U \tag{17}$$

with likelihood function

$$p(Y|\Phi, \Sigma_u) \propto |\Sigma_u|^{-T/2} \exp\left\{ -\frac{1}{2} tr[\Sigma_u^{-1}(Y - X\Phi)'(Y - X\Phi)] \right\} \tag{18}$$

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

3.1 Prior Specification and Posterior

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

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

where θ is the vector of DSGE model parameters. Roughly speaking, the DSGE model prior $p(\Phi, \Sigma_u|\theta)$ is generated by augmenting the actual data with $T^* = \lambda T$ artificial observations generated from the DSGE model. The approach of using artificial or dummy observations to incorporate prior information in VARs is quite common, e.g., Sims and Zha (1998), and originally due to Theil and Goldberger (1961). Rather than generating random observations $y_1^*, \dots, y_{T_*}^*$ from Equation (14) and augmenting the actual data Y, that is, pre-multiplying the likelihood function by

$$\tilde{p}(\Phi, \Sigma_u|\theta) \propto |\Sigma_u|^{-(\lambda T+n+1)/2} \exp\left\{ -\frac{1}{2} tr[\Sigma_u^{-1}(Y^* - X^*\Phi)'(Y^* - X^*\Phi)] \right\}, \tag{19}$$

we will replace the (artificial) sample moments $Y^{*'}Y^*$, $Y^{*'}X^*$, and $X^{*'}X^*$ by population analogs.

If the vector of endogenous variables is composed of output growth, inflation, and the nominal interest rate then y_t is covariance stationary according to the DSGE model (see Section 2). A law of large numbers for weakly dependent processes yields the conclusion that sample moments converge to population moments as the sample size of the dummy observations tends to infinity, e.g.,

$$\lim_{T_* \to \infty} \frac{1}{T_*} Y^{*'} Y^* = \Gamma^*_{yy}(\theta). \tag{20}$$

The limit matrix $\Gamma^*_{yy}(\theta)$ is a function of the structural parameters θ. The probability limits of $Y^{*'}X^*$ and $X^{*'}X^*$ will be denoted by $\Gamma^*_{yx}(\theta)$ and $\Gamma^*_{xx}(\theta)$, respectively.[4]

We now replace the sample moments in Equation (19) by scaled population moments and use the prior

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

The prior is proper provided that $\lambda T \geq k + n$. The proportionality factor $c(\theta)$ ensures that the density integrates to one and is defined in the Appendix. Define the functions

$$\Phi^*(\theta) = \Gamma^{*^{-1}}_{xx}(\theta)\Gamma^*_{xy}(\theta) \tag{22}$$

$$\Sigma^*_u(\theta) = \Gamma^*_{yy}(\theta) - \Gamma^*_{yx}(\theta)\Gamma^{*^{-1}}_{xx}(\theta)\Gamma^*_{xy}(\theta). \tag{23}$$

Conditional on θ the prior distribution of the VAR parameters is of the Inverted-Wishart ($\mathcal{IW}$) – Normal ($\mathcal{N}$) form[5]

$$\Sigma_u|\theta \sim \mathcal{IW}\Big(\lambda T \Sigma^*_u(\theta), \lambda T - k, n\Big) \tag{24}$$

$$\Phi|\Sigma_u, \theta \sim \mathcal{N}\Big(\Phi^*(\theta), \Sigma_u \otimes (\lambda T \Gamma^*_{xx}(\theta))^{-1}\Big). \tag{25}$$

[4]We denote $\Gamma^*_{yx}(\theta)'$ by $\Gamma^*_{xy}(\theta)$.

[5]The construction of the prior is based on the assumption that $\Gamma^*_{xx}(\theta)$ is invertible. This assumption is satisfied in our application as the number of structural shocks equals the number of endogenous variables n to which the model is fitted. In DSGE models with less than n structural shocks the non-singularity of $\Gamma^*_{xx}(\theta)$ could be achieved by the introduction of additional shocks or measurement errors.

Since the likelihood function (18) and prior (21) are conjugate, the posterior distribution is also of the $\mathcal{IW}-\mathcal{N}$ form. While details can be found in the Appendix, it is instructive to examine the posterior mean of Φ conditional on θ. It is given by

$$\tilde{\Phi}(\theta) = \left(\frac{\lambda}{1+\lambda} \Gamma^*_{xx}(\theta) + \frac{1}{1+\lambda} T^{-1} X'X \right)^{-1} \left(\frac{\lambda}{1+\lambda} \Gamma^*_{xy}(\theta) + \frac{1}{1+\lambda} T^{-1} X'Y \right). \tag{26}$$

The hyperparameter λ determines the effective sample size for the artificial observations, which is λT. If $\lambda = 0$ then $\tilde{\Phi}(\theta)$ equals the OLS estimate of Φ. As $\lambda \longrightarrow \infty$, $\tilde{\Phi}(\theta)$ approaches the restriction function $\Phi^*(\theta)$ derived from the DSGE model. For $\lambda = 1$ sample information and prior information receive equal weight in the posterior.

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

$$p_\lambda(Y) = \int p(Y|\Phi, \Sigma_u) p_\lambda(\Phi, \Sigma_u|\theta) p(\theta) d(\Phi, \Sigma_u, \theta) \tag{27}$$

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.

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.

3.2 Interpretation of the Prior

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 $\Sigma_u^*(\theta)$.

The forecast performance of DSGE models is often poor, because they are tightly parameterized and impose some inadequate cross-parameter restrictions on the VAR representation of the data. Therefore, it is important to assign prior probability mass outside of the subspace traced out by $\Phi^*(\theta)$ and $\Sigma_u^*(\theta)$.[6] 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. If λ is large then most of the prior mass concentrates in the vicinity of the subspace $\Phi^*(\theta)$. Our prior is a modification of the one used by DeJong, Ingram, and Whiteman. DeJong, Ingram, and Whiteman 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.

The major improvement of our procedure over earlier approaches is that we compute a joint posterior distribution for Φ, Σ_u, and θ that allows posterior inference with respect to the DSGE model parameters. Since the likelihood function depends on θ only indirectly through Φ and Σ_u the joint posterior can be written as

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

Learning about θ from the data takes place indirectly through learning about the VAR parameters Φ, Σ_u. The information on θ will play an important role for the policy analysis in Section 4.

[6] Ingram and Whiteman 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.

3.3 Learning about the DSGE Model Parameters

The purpose of this section is to characterize some of the properties of the posterior distribution of the DSGE model parameters. The post-sample information about the DSGE model parameters is summarized in the marginal posterior density of θ

$$p(\theta|Y) = \int p(\theta|\Phi, \Sigma_u) p(\Phi, \Sigma_u|Y) d(\Phi, \Sigma_u). \tag{29}$$

Under the improper prior $\lambda = 0$ the VAR parameters and the DSGE model parameters are *a priori* independent. Since the likelihood function does not depend on θ, its prior is not updated and nothing is learnt about the structural parameters. However, if $\lambda > 0$ then Φ, Σ_u, and θ are correlated *a priori* and the data become informative about the structural parameters. Roughly speaking, the conditional prior density $p(\theta|\Phi, \Sigma_u)$ projects the posterior estimates of the VAR parameters back onto the space traced out by $\Phi^*(\theta)$ and $\Sigma_u^*(\theta)$ and its mode can be interpreted as minimum-distance estimator of the DSGE model parameters. This claim is subsequently formalized through two asymptotic approximations.

Let us start by defining the quasi-likelihood function[7]

$$p^*(Y|\theta) \propto |\Sigma_u^*(\theta)|^{-T/2} \exp\left\{ -\frac{1}{2} tr \left[\Sigma_u^{*^{-1}}(\theta)(Y - X\Phi^*(\theta))'(Y - X\Phi^*(\theta)) \right] \right\}. \tag{30}$$

The logarithm of the quasi-likelihood function can be approximated as follows[8]

$$\begin{aligned} & \ln p^*(Y|\theta) \qquad (31) \\ \approx \; & const - \frac{1}{2} vec(\hat{\Phi}_{mle} - \Phi^*(\theta))'(\hat{\Sigma}_{u,mle}^{-1} \otimes X'X) vec(\hat{\Phi}_{mle} - \Phi^*(\theta)) \\ & - \frac{T}{2} vech(\hat{\Sigma}_{u,mle} - \Sigma_u^*(\theta))' D (\hat{\Sigma}_{u,mle}^{-1} \otimes \hat{\Sigma}_{u,mle}) D' vech(\hat{\Sigma}_{u,mle} - \Sigma_u^*(\theta))', \end{aligned}$$

where $\hat{\Phi}_{mle} = (X'X)^{-1}X'Y$ and $\hat{\Sigma}_{u,mle} = (Y'Y - Y'X(X'X)^{-1}X'Y)/T$ are the maximum-likelihood estimators of the VAR parameters Φ and Σ_u. The *vec*-operator stacks the columns of a matrix, the *vech*-operator stacks the non-redundant elements

[7] Since 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.

[8] The approximation is obtained from a second-order Taylor expansion of $p^*(Y|\theta)$ around $\Phi^* = \hat{\Phi}_{mle}$ and $\Sigma_u^* = \hat{\Sigma}_{u,mle}$.

of a symmetric matrix, and D is the duplicator matrix that satisfies $vec(A) = Dvech(A)$. The direct maximization of the quasi-likelihood function $p^*(Y|\theta)$ with respect to the DSGE model parameters θ can be regarded as minimum distance estimation (e.g., Chamberlain (1984) and Moon and Schorfheide (2002)) as θ is essentially estimated by minimizing the weighted discrepancy between the unrestricted VAR estimates $\hat{\Phi}_{mle}$ and the restriction function $\Phi^*(\theta)$.

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

$$p(Y|\theta) = \int p(Y|\Phi, \Sigma_u)p(\Phi, \Sigma_u|\theta)d(\Phi, \Sigma_u) \tag{32}$$

with the prior $p(\theta)$. Suppose the sample size T is fixed and the number of artificial observations from the DSGE model is large. As $\lambda \longrightarrow \infty$ the prior for the VAR parameters concentrates its mass near the subspace traced out by $\Phi^*(\theta)$ and $\Sigma_u^*(\theta)$. Hence, the marginal likelihood function $p(Y|\theta)$ is approximately equal to the quasi-likelihood function $p^*(Y|\theta)$, defined in Equation (30).

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

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

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

Based on Proposition 1 and Equation (31) we can deduce that the posterior mode of the marginal log-likelihood function approximately minimizes the weighted discrepancy between the VAR estimates $\hat{\Phi}_{mle}$, $\hat{\Sigma}_{u,mle}$ and the restriction functions $\Phi^*(\theta)$, $\Sigma_u^*(\theta)$.

For reasons discussed in Section 3.2 we expect that the best fit of the vector autoregression model is achieved for moderate values of λ. Hence, we consider a second approximation in which the sample size is large ($T \longrightarrow \infty$), yet the relative importance of the prior is modest ($\lambda \longrightarrow 0$, $\lambda T \longrightarrow \infty$). Define the function

$$q(\theta|Y) \quad = \quad \exp\left\{-\frac{1}{2}\ln|\Sigma_u^{*^{-1}}(\theta)| - \frac{1}{2}tr[\hat{\Sigma}_{u,mle}^{-1}\Sigma_u^*(\theta)]\right. \tag{33}$$

$$-\frac{1}{2}tr[\hat{\Sigma}_{u,mle}^{-1}(\Phi^*(\theta)-\hat{\Phi}_{mle})'\Gamma_{xx}^*(\theta)(\Phi^*(\theta)-\hat{\Phi}_{mle})]\bigg\}.$$

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, see Equation (31).[9] The marginal log-likelihood function can be approximated as follows.

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

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,

[9] The weights are ${\Sigma_u^*}^{-1}\otimes\Gamma_{xx}^*(\theta)$ and $D'({\Sigma_u^*}^{-1}\otimes{\Sigma_u^*}^{-1})D$ instead of $\hat{\Sigma}_{u,mle}^{-1}\otimes\hat{\Gamma}_{xx}$ and $D'(\hat{\Sigma}_{u,mle}^{-1}\otimes\hat{\Sigma}_{u,mle}^{-1})D$.

inflation, and interest rates.[10] Section 4 consists of four parts. Section 4.1 discusses the prior and posterior for the DSGE model parameters. Section 4.2 describes the forecasting results. There, we show that the introduction of the prior produces a substantial improvement relative to an unrestricted VAR in terms of forecasting. We also show that the forecasting performance of the VAR with DSGE priors (VAR-DSGE) is competitive relative to that of a VAR with Minnesota priors (VAR-Minn).

In Sections 4.3 and 4.4 we describe how policy analysis can be conducted using a VAR with DSGE model prior. We consider two types of experiments. The first experiment is what Leeper and Zha (2001) call a "modest policy intervention": a very short sequence of small policy shocks. Researchers typically conduct these experiments using impulse response functions that come from either DSGE models or "identified" vector autoregressions. The key to this type of experiments is therefore the identification of monetary policy shocks. The identification approach we propose follows naturally from the overall strategy of the paper: we use the VAR approximation of the DSGE model's impulse responses as a prior for the VAR impulse responses.

The second type of policy experiment consists of forecasting the effects of a policy rule change. Due to the Lucas' critique, this kind of experiment is generally considered infeasible within the identified VAR framework. In this sense, the second experiment is more ambitious than the first. As a rough approximation, our approach can be seen as a weighted average between 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

[10] We use quarterly data. The data for real output growth come from the Bureau of Economic Analysis (Gross Domestic Product-SAAR, Billions Chained 1996$). The data for inflation come from the Bureau of Labor Statistics (CPI-U: All Items, seasonally adjusted, 1982-84=100). The interest rate series are constructed as in Clarida, Galí, and Gertler (2000): for each quarter the interest rate is computed as the average federal funds rate (source: Haver Analytics) during the first month of the quarter, including business days only. The data are available from 1955:III to 2001:III. The lag length in the VAR is four quarters.

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 VAR-DSGE. 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 percentage points. 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 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 1979:III to 1999: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 and the slope of the expectational Phillips curve and the response of the central bank to output are revised upwards. The updating is more pronounced for $\lambda = 10$, when the artificial sample size is ten times as long as the actual sample. The empirical findings are consistent with Proposition 2 which implies that information about θ is accumulated at rate λT.

[11] Since 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.

For the empirical results an appropriate choice of the weight of the prior λ is very important. We argued in Section 3.1 that λ can be chosen over a grid Λ to maximize the marginal data density $p_\lambda(Y)$, given in Equation (27). Below we will often refer to the hyperparameter estimate

$$\hat{\lambda} = \text{argmax}_{\lambda \in \Lambda}\, p_\lambda(Y). \tag{34}$$

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 0.4 and 1, suggesting that the fit of the model is roughly the same within that range.[12]

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 endogenous variables.

In a particular instance, this point has already been made by Ingram and Whiteman. 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, unlike Ingram and Whiteman, we select the relative weight of the prior *ex ante*, based on the marginal posterior density of the hyperparameter λ. 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.

[12] A 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.

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). The optimal weight $\hat{\lambda}$ is computed for each forecasting origin of the rolling sample. 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. 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.

How does the forecasting performance of VAR-DSGE rank relative to an unrestricted VAR, and a VAR-Minn? Table 3 provides the percentage improvement (or loss, if negative) in root mean square forecast errors (*rmse*) of VAR-DSGE relative to both competitors for cumulative real output growth, cumulative inflation, and the federal funds rate. The improvement in *rmse*s is shown for one, two, four, six, eight, ten, twelve, and sixteen quarters ahead.[13] Table 3 also reports the improvements in the multivariate forecasting performance statistic proposed by Doan, Litterman, and Sims (1984).[14]

Let us first focus on the comparison with the unrestricted VAR. Our results indicate VAR-DSGE performs better than the unrestricted VAR for all variables at all horizons. Quantitatively, the improvements are large for all variables. In terms of the multivariate statistics the improvements range from a minimum of 9% for four quarters ahead forecasts, to a maximum of almost 12% for forecasts four years

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

[14]The *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.

ahead. The improvement increases almost monotonically with the forecast horizon for all variables except the federal funds rate.

When forecasters use prior information in VARs, they mostly use Minnesota priors. Hence, VAR-Minn is a natural competitor with our approach.[15] Table 3 shows that VAR-Minn performs better than VAR-DSGE for very short run forecasts, especially for the federal funds rate and real output growth. As the forecast horizon increases the relative performance of VAR-DSGE improves: for horizons beyond one quarter the multivariate statistics suggest that VAR-DSGE outperforms VAR-Minn. The improvement is sizable for output and inflation forecast (up to 20 and 7 %, respectively), and is non-existent for federal funds forecasts. Overall, these results suggest that VAR-DSGE is competitive with, and sometimes improves upon, VAR-Minn for forecasts beyond the very short run. When interpreting these results one has to bear in mind a key difference between Minnesota and DSGE prior. The Minnesota prior has a statistical (unit root processes fit a number of economic series quite well) but not necessarily an economic justification. DSGE priors do: We know where they come from. We know how to interpret them. In addition, Minnesota priors may help to forecast well in some dimensions, but offer no help when it comes to policy analysis. This is not the case for DSGE priors, as discussed in Sections 4.3 and 4.4.

Next we investigate how the forecasting performance of the VAR changes as a

[15]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 (27). This value hovers around 0.5, depending on the sample. The value used in Doan, Litterman and Sims (1984) is $\iota = 1$.

function of λ. Figure 1 plots the percentage improvement relative to the unrestricted VAR for a grid of values of λ ranging from 0 to ∞ ($\lambda = \infty$ means forecasting with the VAR approximation to the DSGE model). By definition the gain for $\lambda = 0$ is zero. The one-step ahead forecasting performance peaks around $\lambda = 0.8$, which is roughly consistent with the weight selected *ex ante* based on the marginal data density $p_\lambda(Y)$. The short run forecasting performance remains competitive even for relatively large values of λ ($\lambda = 5$), but deteriorates substantially for $\lambda = \infty$, especially for forecasts of real output growth. For medium and long-run forecast horizons the best multivariate forecasting performance is achieved for a value of λ of approximately 2. 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 VAR-DSGE are still accurate even when the prior weight is infinity.

In summary, this section shows that the VAR with DSGE prior is a fairly competitive model in terms of forecasting. VAR-DSGE is inferior to VAR-Minn for one quarter ahead forecasts, but otherwise holds its own and often outperforms its competitors, sometimes by sizable margins.[16] The section also shows that relying on the DSGE model only for forecasting ($\lambda = \infty$) can lead to imprecise forecasts, especially in the short run.

[16]We 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 VAR-DSGE and VAR-Minn *ex ante*.

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 map the one-step-ahead forecast errors u_t into the structural shocks ϵ_t. Let Σ_u^C 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:

$$u_t = \Sigma_u^C \Omega \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \mathcal{I}_n), \tag{35}$$

where Ω is an orthonormal matrix. The identification problem arises from the fact that the data are silent about the choice of the rotation matrix Ω. More prosaically, since $\Sigma_u^C \Omega \Omega' \Sigma_u^{C'} = \Sigma_u^C \Sigma_u^{C'}$ 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. For example, Blanchard and Quah (1989) focus on the long-run properties of shocks, while Faust (1998), Canova and DeNicoló (2001) and Uhlig (2001) focus on sign restrictions on impact.

In our identification strategy the theoretical model serves as a prior for the VAR impulse responses, which is consistent with the overall approach of the paper.[17] The extent to which the posterior impulse responses are forced to look like the model's responses will depend on the tightness of the prior. Our procedure has two main advantages. First, once the theoretical model is chosen there is no room for arbitrariness. Conditional on the weight of the prior, the data – and not 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. Second,

[17] It is noteworthy that in principle the issue of identification is completely decoupled from that of forecasting: one could use any of the available approaches to identification in VARs, e.g., linear restrictions on the covariance matrix of the innovations as in Bernanke (1986) and Sims (1986), and still use DSGE priors for the VAR parameters.

whether the model can be considered a reliable basis for identification is determined by its fit, as the tightness of the prior is chosen endogenously.

The DSGE model is of course identified: For each set of deep parameters θ, there is a unique matrix $\Omega^*(\theta)$ that maps the Cholesky decomposition of the variance-covariance matrix of forecast errors into the matrix of the DSGE impulse responses on impact. The computation of $\Omega^*(\theta)$ is straightforward. If $A_0(\theta)$ is the matrix of DSGE impulse responses on impact obtained from Equation (14), the QR decomposition of $A_0(\theta)$ – available in most computer packages – will deliver a lower triangular matrix $\Sigma^{C*}_{DSGE}(\theta)$ and a unitary matrix $\Omega^*(\theta)$ such that $A_0(\theta) = \Sigma^{C*}_{DSGE}(\theta) * \Omega^*(\theta)$. Let us call the triplet (Φ, Σ_u, Ω) the parameters of the identified VAR. Through the identified VAR approximation of the DSGE model, given by $(\Phi^*(\theta), \Sigma^*_u(\theta), \Omega^*(\theta))$, the prior distribution of the DSGE model parameters θ induces a prior distribution for the identified VAR parameters.[18]

The posterior distribution is obtained by updating the distribution of Φ, Σ_u, and θ as described in Section 3 and mapping θ into $\Omega = \Omega^*(\theta)$. Conditional on θ, the rotation matrix is the same *a posteriori* as it is *a priori*, since the likelihood function of the reduced form VAR is invariant with respect to Ω. However, we learn from the data which rotation to choose, albeit indirectly, via learning about the DSGE model parameters θ. Moreover, even conditional on θ, the posterior VAR impulse responses will differ from the prior responses, to the extent that the distribution of Φ and Σ_u is being updated.

There are a few attempts in the literature to parameterize the VAR in terms of its moving-average (MA) representation and to specify a prior distribution directly for the impulse responses, subject to some restrictions that ensure that the MA representation is consistent with a finite-order VAR, e.g., Dwyer (1998) and Gordon and Boccanfuso (2001). However, the difference to the approach proposed in our paper lies merely in the construction of the prior distribution. No matter how such

[18] Since 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 $\Sigma^{C*}_{DSGE}(\theta)$. However, in our experience the difference is for practical purposes negligible.

a prior is constructed, the likelihood function is always flat in directions of the parameter space in which Φ and Σ_u are constant and the conditional distribution of Ω given Φ and Σ_u is never updated. Our parameterization of the VAR simply makes explicit in which directions of the parameter space learning from the data is possible.

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 plots 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. Also, the responses of inflation to a money shock is short-lived both in the 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 (2001), 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. This section discusses how to use a VAR with DSGE model prior to analyze the effects of regime shifts. The joint posterior distribution for the VAR and DSGE model parameters can be decomposed into

$$p(\Phi, \Sigma_u, \theta|Y) = p(\Phi, \Sigma_u|\theta, Y)p(\theta|Y). \tag{36}$$

Assessing the effects of a policy regime shift is equivalent to the modification of the posterior $p(\theta|Y)$. Partition $\theta = [\theta'_{(s)}, \theta'_{(p)}]'$, where $\theta_{(p)}$ corresponds to the policy parameters that are affected by the regime shift. We construct a modified posterior

$$\tilde{p}(\theta|Y) = p(\theta_{(s)})\tilde{p}(\theta_{(p)}), \tag{37}$$

where $p(\theta_{(s)}|Y)$ is the marginal posterior of the non-policy parameters and $\tilde{p}(\theta_{(p)})$ is a (possibly degenerate) distribution that determines the value of the policy parameters in the experiment.

To translate the structural parameters back into VAR parameters, we will use $p(\Phi, \Sigma_u|\theta, Y)$. For the sake of concreteness, let us recall that the posterior mean of Φ conditional on θ is given by the formula:

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

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

$$\tilde{p}_\lambda(\Phi, \Sigma_u|Y) = \int p(\Phi, \Sigma_u|Y, \theta)\tilde{p}(\theta|Y)d\theta. \tag{38}$$

We use the subscript λ to indicate that the conclusions depend on the weight given to the DSGE model. 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. In all other cases the forecasts are partially based on data that are generated by an economy with 'old' policy parameters. The policy analyst may be willing to pay this price – which is not necessarily high for high values of λ – if she can reap substantial gains in forecasting performance.

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 (13)) 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 policy[19], 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 (37) 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

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

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.[20]

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 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. The predicted effect of the policy change becomes larger as the weight of the prior increases. Moreover, the uncertainty about the variability decreases. The predictions for the standard deviation of output do not change as ψ_1 increases.

Figure 3 also shows that the two extremes ($\lambda = 0$ and $\lambda = \infty$) seem to lead to inaccurate evaluations of the effects of the policy. The analyst who completely ignores the policy change ($\lambda = 0$) grossly over-predicts the standard deviation of inflation. The analyst who relies only on the DSGE model ($\lambda = \infty$) under-predicts the standard deviation of inflation, albeit to a lesser extent, and the standard deviation of the federal funds rate. The marginal data density $p_\lambda(Y)$ suggests *ex ante* to avoid the extremes and to choose a λ between 0.5 and 1. Values between 1

[20] A 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).

and 5 seem to deliver the most accurate predictions *ex post*. VAR-DSGE 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.

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. One can interpret the density $p_\lambda(\Phi, \Sigma_u|\theta, Y)$ as a "correction" to the vector autoregressive representation of the DSGE model given the structural parameters θ. This "correction" has been constructed from past observations to optimize the forecasting performance. Our approach is based on the presumption that in the absence of contrary evidence it is reasonable to proceed as if the "correction" is policy invariant.[21] Whenever misspecified models are used for policy analysis it is typically assumed that the misspecification is policy invariant and that pre-intervention corrections remain valid in the new regime.

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 strategy, usable for policy analysis. The strategy 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

[21] Determining *a priori* whether or not this invariance is satisfied in practice is infeasible. In fact, it would require *ex ante* knowledge about the actual effects of the policy, which is typically not available. If it were available, the DSGE model based policy analysis would not be interesting.

posterior distribution of DSGE and VAR parameters, which can then be used to compute predictive densities. The strategy is very efficient from the computational point of view.

We apply the strategy to a VAR in real output growth, inflation, and interest rates, and show that it is broadly successful in terms of forecasting performance. The VAR with DSGE model prior clearly outperforms an unrestricted VAR at all horizons. Its forecasting performance is comparable to a VAR with Minnesota prior. While the Minnesota prior is helpful for forecasting, it offers no help when it comes to policy analysis.

We provide an identification scheme for the structural shocks and hence enable an analysis of modest policy interventions. Our approach follows naturally from the overall strategy in the paper. Construct an orthonormal matrix from the VAR approximation of the DSGE model to map the reduced form innovations into structural shocks. This orthonormal matrix induces a DSGE model based prior distribution for VAR impulse responses that can be updated with the available data. We argue that our approach produces an attractive alternative to existing identification schemes – attractive because it ties identification to a fully specified general equilibrium model, leaving no room for arbitrariness.

We also illustrate how a VAR with DSGE model prior can be used to predict the effects of changes in the policy regime – 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 better than using the DSGE model only, or the unrestricted VAR only, to predict the effect of the change, although further research is needed to investigate this issue more deeply.

As envisioned in Diebold (1998), the combination of DSGE models and vector autoregressions shows promise for macroeconomic forecasting and policy analysis. Yet, more research lies down the road. If the VAR is specified in terms of output and prices rather than output growth and inflation, then the asymptotic behavior of the sample moments of the artificial data changes. The elements of the properly standardized moment matrices have stochastic rather than deterministic limits and our construction of the prior has to be modified. Moreover, it is worthwhile 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.

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, **25**, Amsterdam: North-Holland, 49-99.

Blanchard, Olivier Jean and Danny Quah (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances". *American Economic Review*, **79**, 655-673.

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

Chamberlain, Gary (1984): "Panel Data," in Z. Griliches and M. Intriligator, eds. *Handbook of Econometrics*, Vol. 3, New York, 1465-1514.

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.

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.

Galí, Jordi and Mark Gertler (1999): "Inflation Dynamics: A Structural Econometric Analysis". *Journal of Monetary Economics*, **44**(2), pp. 195-222.

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

Gordon, Stephen and Dorothea Boccanfuso (2001) "Learning from Structural Vector Autoregression Models," *Manuscript*, Universite Laval, Quebec City.

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. (1999) "A Method for Taking Models to the Data." *Manuscript*, Boston College.

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 (2001) "Modest Policy Interventions." *Federal Reserve Bank of Atlanta Working Paper*, **99-22**.

Lubik, Thomas and Frank Schorfheide (2002): "Testing for Indeterminacy in Linear Rational Expectations Models". *Manuscript*, Department of Economics, University of Pennsylvania.

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

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.

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

——————— (2000): "Solving Rational Expectations Models". *Manuscript*, Department of Economics, Princeton University.

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

Smets, Frank and Rafael Wouters (2002): "An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area". *Manuscript*, European Central Bank.

Taylor, John B. (1999): "*Monetary Policy Rules*". University of Chicago Press.

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 (2000): "A Neo-Wicksellian Framework for the Analysis of Monetary Policy". *Manuscript*, Department of Economics Princeton University.

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

A Analysis of the Posterior Distribution

Prior Distribution: Conditional on the DSGE model parameters θ the prior density for the VAR parameters is of the form

$$\begin{aligned} p(\Phi, \Sigma_u|\theta) &= c^{-1}(\theta)|\Sigma_u|^{-\frac{\lambda T+n+1}{2}} \\ &\times \exp\left\{-\frac{1}{2}tr[\lambda T\Sigma_u^{-1}(\Gamma^*_{yy}(\theta) - \Phi'\Gamma^{*\prime}_{yx}(\theta) - \Gamma^*_{xy}(\theta)\Phi + \Phi'\Gamma^*_{xx}(\theta)\Phi)]\right\}. \end{aligned} \tag{A1}$$

The normalization factor $c(\theta)$ is

$$\begin{aligned} c(\theta) &= (2\pi)^{\frac{nk}{2}}|\lambda T\Gamma^*_{xx}(\theta)|^{-\frac{n}{2}}|\lambda T\Sigma^*_u(\theta)|^{-\frac{\lambda T-k}{2}} \\ & 2^{\frac{n(\lambda T-k)}{2}}\pi^{\frac{n(n-1)}{4}}\prod_{i=1}^{n}\Gamma[(\lambda T-k+1-i)/2], \end{aligned} \tag{A2}$$

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

Posterior Distribution: In order to analyze the posterior distribution we use the following factorization

$$p(\Phi, \Sigma_u, \theta|Y) = p(\Phi, \Sigma_u|Y, \theta)p(\theta|Y). \tag{A3}$$

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

$$\begin{aligned} \tilde{\Phi}(\theta) &= (\lambda T \Gamma^*_{xx}(\theta) + X'X)^{-1}(\lambda T \Gamma^*_{xy} + X'Y) && \text{(A4)} \\ \tilde{\Sigma}_u(\theta) &= \frac{1}{(\lambda+1)T}\Big[(\lambda T \Gamma^*_{yy}(\theta) + Y'Y) \\ &\quad -(\lambda T \Gamma^*_{yx}(\theta) + Y'X)(\lambda T \Gamma^*_{xx}(\theta) + X'X)^{-1}(\lambda T \Gamma^*_{xy}(\theta) + X'Y)\Big]. && \text{(A5)} \end{aligned}$$

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:

$$\begin{aligned} \Sigma_u|Y, \theta &\sim \mathcal{IW}\Big((\lambda+1)T\tilde{\Sigma}_u(\theta), (1+\lambda)T - k, n\Big) && \text{(A6)} \\ \Phi|Y, \Sigma_u, \theta &\sim \mathcal{N}\Big(\tilde{\Phi}(\theta), \Sigma_u \otimes (\lambda T \Gamma^*_{xx}(\theta) + X'X)^{-1}\Big). && \text{(A7)} \end{aligned}$$

The marginal likelihood function of θ is given by

$$\begin{aligned} p(Y|\theta) &= \int p(Y|\Phi, \Sigma_u)p(\Phi, \Sigma_u|\theta)d(\Phi, \Sigma_u) && \text{(A8)} \\ &= \frac{p(Y|\Phi, \Sigma)p(\Phi, \Sigma|\theta)}{p(\Phi, \Sigma|Y)} \\ &= \frac{|\lambda T \Gamma^*_{xx}(\theta) + X'X|^{-\frac{n}{2}}|(\lambda+1)T\tilde{\Sigma}_u(\theta)|^{-\frac{(\lambda+1)T-k}{2}}}{|\lambda T \Gamma^*_{xx}(\theta)|^{-\frac{n}{2}}|\lambda T \Sigma^*_u(\theta)|^{-\frac{\lambda T - k}{2}}} \\ &\quad \times \frac{(2\pi)^{-nT/2} 2^{\frac{n((\lambda+1)T-k)}{2}} \prod_{i=1}^n \Gamma[((\lambda+1)T - k + 1 - i)/2]}{2^{\frac{n(\lambda T - k)}{2}} \prod_{i=1}^n \Gamma[(\lambda T - k + 1 - i)/2]}. \end{aligned}$$

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

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p_\lambda(Y)}, \tag{A9}$$

where

$$p_\lambda(Y) = \int p(Y|\theta)p(\theta)d\theta \tag{A10}$$

is the marginal data density, indexed by the hyperparameter λ.

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

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

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

(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 standard methods to generate draws from $p(\Phi, \Sigma_u|Y, \theta_{(s)})$ 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.

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 (A8). Consider the log-likelihood ratio

$$\begin{aligned} \ln \frac{p(Y|\theta)}{p(Y|\tilde{\theta})} &= -\frac{T}{2} \ln |\Sigma_u^*(\theta)| - \frac{n}{2} \ln |\mathcal{I} + \phi \Gamma_{xx}^{*^{-1}}(\theta)\hat{\Gamma}_{xx}| \\ &\quad - \frac{(1/\phi + 1)T - k}{2} \ln \left| \frac{1/\phi + 1}{1/\phi} \Sigma_u^{*^{-1}}(\theta)\tilde{\Sigma}_u(\theta) \right| \\ &\quad + \frac{T}{2} \ln |\Sigma_u^*(\tilde{\theta})| + \frac{n}{2} \ln |\mathcal{I} + \phi \Gamma_{xx}^{*^{-1}}(\tilde{\theta})\hat{\Gamma}_{xx}| \\ &\quad + \frac{(1/\phi + 1)T - k}{2} \ln \left| \frac{1/\phi + 1}{1/\phi} \Sigma_u^{*^{-1}}(\tilde{\theta})\tilde{\Sigma}_u(\tilde{\theta}) \right|. \end{aligned} \tag{A11}$$

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

$$\ln |\mathcal{I} + \phi \Gamma_{xx}^{*^{-1}} \hat{\Gamma}_{xx}| = \ln |\mathcal{I}| + \phi tr[\Gamma_{xx}^{*^{-1}} \hat{\Gamma}_{xx}| + O(\phi^2). \tag{A12}$$

Notice that

$$\frac{1/\phi+1}{1/\phi}\Sigma_u^{*^{-1}}\tilde{\Sigma}_u = \left[\Gamma^*_{yy} - \Gamma^*_{yx}\Gamma^{*^{-1}}_{xx}\Gamma^*_{xy}\right]^{-1} \times\left[\Gamma^*_{yy} + \phi\hat{\Gamma}_{yy} - (\Gamma^*_{yx}+\phi\hat{\Gamma}_{yx})(\Gamma^*_{xx}+\phi\hat{\Gamma}_{xx})^{-1}(\Gamma^*_{xy}+\phi\hat{\Gamma}_{xy})\right]. \tag{A13}$$

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

$$\ln\left|\frac{1/\phi+1}{1/\phi}\Sigma_u^{*^{-1}}\tilde{\Sigma}_u\right| = \ln|\mathcal{I}| + \phi tr\left[\Sigma_u^{*^{-1}}(\hat{\Gamma}_{yy} - \hat{\Gamma}_{yx}\Phi^* - \Phi^{*'}\hat{\Gamma}_{xy} + \Phi^{*'}\hat{\Gamma}_{xx}\Phi^*)\right] + O(\phi^2). \tag{A14}$$

Combining these results yields

$$\begin{aligned}\ln p(Y|\theta) = & -\frac{T}{2}\ln|\Sigma^*_u(\theta)| + \frac{T}{2}\ln|\Sigma^*_u(\tilde{\theta})| \\ & -\frac{T}{2}tr\left[\Sigma_u^{*^{-1}}(\theta)(\hat{\Gamma}_{yy} - \hat{\Gamma}_{yx}\Phi^*(\theta) - \Phi^{*'}(\theta)\hat{\Gamma}_{xy} + \Phi^{*'}(\theta)\hat{\Gamma}_{xx}\Phi^*(\theta))\right] \\ & +\frac{T}{2}tr\left[\Sigma_u^{*^{-1}}(\tilde{\theta})(\hat{\Gamma}_{yy} - \hat{\Gamma}_{yx}\Phi^*(\tilde{\theta}) - \Phi^{*'}(\tilde{\theta})\hat{\Gamma}_{xy} + \Phi^{*'}(\tilde{\theta})\hat{\Gamma}_{xx}\Phi^*(\tilde{\theta}))\right] + O(\phi) \\ = & \ln\frac{p^*(Y|\theta)}{p^*(Y|\tilde{\theta})} + O(\phi).\end{aligned} \tag{A15}$$

Thus, as $\phi \longrightarrow 0$ 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 (A11) in terms of λ:

$$\begin{aligned}\ln\frac{p(Y|\theta)}{p(Y|\tilde{\theta})} = & -\frac{T}{2}\ln|\Sigma^*_u(\theta)| - \frac{n}{2}\ln|\lambda\mathcal{I} + \Gamma^{*^{-1}}_{xx}(\theta)\hat{\Gamma}_{xx}| \\ & -\frac{(\lambda+1)T-k}{2}\ln\left|(\lambda+1)\Sigma_u^{*^{-1}}(\theta)\tilde{\Sigma}_u(\theta)\right| \\ & +\frac{T}{2}\ln|\Sigma^*_u(\tilde{\theta})| + \frac{n}{2}\ln|\lambda\mathcal{I} + \Gamma^{*^{-1}}_{xx}(\tilde{\theta})\hat{\Gamma}_{xx}| \\ & +\frac{(\lambda+1)T-k}{2}\ln\left|(\lambda+1)\Sigma_u^{*^{-1}}(\tilde{\theta})\tilde{\Sigma}_u(\tilde{\theta})\right|.\end{aligned} \tag{A16}$$

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

$$\ln|\lambda\mathcal{I} + \Gamma^{*^{-1}}_{xx}\hat{\Gamma}_{xx}| = \ln|\Gamma^{*^{-1}}_{xx}\hat{\Gamma}_{xx}| + \lambda tr[\hat{\Gamma}^{-1}_{xx}\Gamma^*_{xx}] + O(\lambda^2). \tag{A17}$$

Notice that

$$(\lambda+1)\Sigma_u^{*^{-1}}\tilde{\Sigma}_u = \left[\Gamma^*_{yy} - \Gamma^*_{yx}\Gamma^{*^{-1}}_{xx}\Gamma^*_{xy}\right]^{-1} \times\left[\lambda\Gamma^*_{yy} + \hat{\Gamma}_{yy} - (\lambda\Gamma^*_{yx} + \hat{\Gamma}_{yx})(\lambda\Gamma^*_{xx} + \hat{\Gamma}_{xx})^{-1}(\lambda\Gamma^*_{xy} + \hat{\Gamma}_{xy})\right]. \quad \text{(A18)}$$

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

$$\begin{aligned} &\ln\left|(\lambda+1)\Sigma_u^{*^{-1}}\tilde{\Sigma}_u\right| \quad \text{(A19)} \\ &= \ln|\Sigma_u^{*^{-1}}\hat{\Sigma}_{u,mle}| + \lambda tr\left[\hat{\Sigma}^{-1}_{u,mle}(\Gamma^*_{yy} - \Gamma^*_{yx}\hat{\Phi}_{mle} - \hat{\Phi}'_{mle}\Gamma^*_{xy} + \hat{\Phi}'_{mle}\Gamma^*_{xx}\hat{\Phi}_{mle})\right] + O(\lambda^2) \\ &= \ln|\Sigma_u^{*^{-1}}\hat{\Sigma}_{u,mle}| + \lambda tr\left[\hat{\Sigma}^{-1}_{u,mle}\Sigma^*_u\right] + \lambda tr\left[\hat{\Sigma}^{-1}_{u,mle}(\Phi^* - \hat{\Phi}_{mle})'\Gamma^*_{xx}(\Phi^* - \hat{\Phi}_{mle})\right] + O(\lambda^2). \end{aligned}$$

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

$$\begin{aligned} \frac{p(Y|\theta)}{p(Y|\tilde{\theta})} &= -\frac{\lambda T}{2}\ln|\Sigma_u^{*^{-1}}(\theta)| + \frac{\lambda T}{2}\ln|\Sigma_u^{*^{-1}}(\tilde{\theta})| \quad \text{(A20)} \\ &\quad -\frac{\lambda T}{2}tr[\hat{\Sigma}^{-1}_{u,mle}\Sigma^*_u(\theta)] + \frac{\lambda T}{2}tr[\hat{\Sigma}^{-1}_{u,mle}\Sigma^*_u(\tilde{\theta})] \\ &\quad -\frac{\lambda T}{2}tr[\hat{\Sigma}^{-1}_{u,mle}(\Phi^*(\theta) - \hat{\Phi}_{mle})'\Gamma^*_{xx}(\theta)(\Phi^*(\theta) - \hat{\Phi}_{mle})] \\ &\quad +\frac{\lambda T}{2}tr[\hat{\Sigma}^{-1}_{u,mle}(\Phi^*(\tilde{\theta}) - \hat{\Phi}_{mle})'\Gamma^*_{xx}(\tilde{\theta})(\Phi^*(\tilde{\theta}) - \hat{\Phi}_{mle})] + O_p(\max[\lambda^2 T, 1]) \\ &= \ln\frac{q(Y|\theta)}{q(Y|\tilde{\theta})} + O_p(\max[\lambda^2 T, 1]). \end{aligned}$$

Table 1: Prior Distributions for DSGE Model Parameters

Name	Range	Density	Mean	S.D.
$\ln\gamma$	$I\!R$	Normal	0.500	0.250
$\ln\pi^*$	$I\!R$	Normal	1.000	0.500
$\ln r^*$	$I\!R^+$	Gamma	0.500	0.250
κ	$I\!R^+$	Gamma	0.300	0.150
τ	$I\!R^+$	Gamma	2.000	0.500
ψ_1	$I\!R^+$	Gamma	1.500	0.250
ψ_2	$I\!R^+$	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	$I\!R^+$	Inv. Gamma	0.251	0.139
σ_g	$I\!R^+$	Inv. Gamma	0.630	0.323
σ_z	$I\!R^+$	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 percentage points. 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: 1979:III - 1999: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	0.438	0.885	0.553	0.846
$\ln \pi^*$	0.219	1.863	0.505	1.465	0.415	1.310
$\ln r^*$	0.132	0.880	0.272	0.967	0.560	0.969
κ	0.063	0.513	0.302	0.918	0.398	0.896
τ	1.197	2.788	0.716	1.816	0.667	1.585
ψ_1	1.121	1.910	1.133	1.810	1.476	2.077
ψ_2	0.001	0.260	0.092	0.501	0.082	0.330
ρ_R	0.157	0.812	0.211	0.536	0.426	0.612

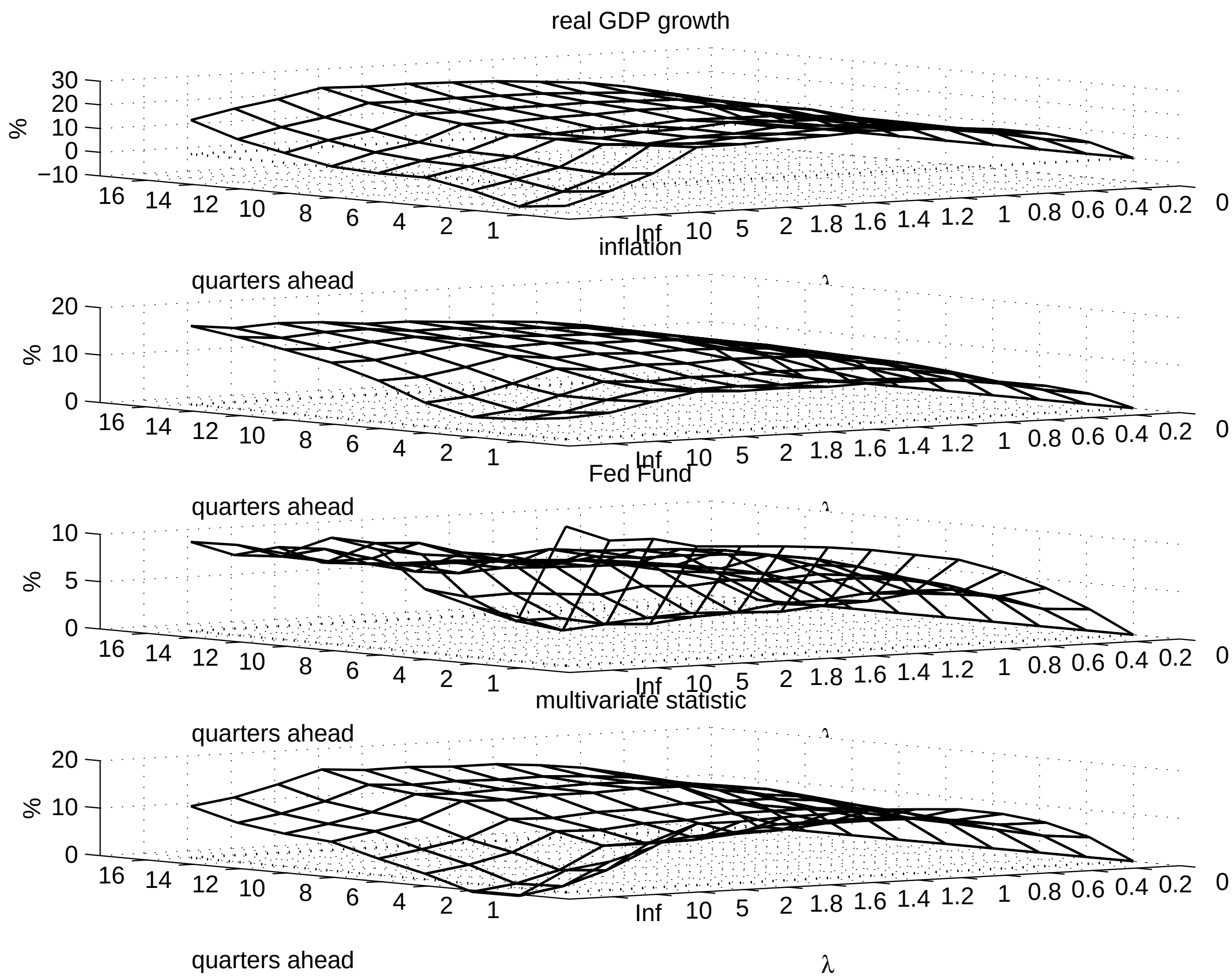
Notes: We report 90 % confidence intervals based on the output of the Metropolis-Hastings Algorithm. The model parameters $\ln \gamma$, $\ln \pi^*$, and $\ln r^*$ are scaled by 100 to converted them into percentage points.

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	15.000	-1.721	6.630	-0.235	7.338	-7.491	11.241	-0.658
2	13.490	3.057	6.367	0.403	4.785	-5.158	9.049	0.940
4	12.986	3.505	7.736	3.697	4.821	-2.078	8.767	3.096
6	13.102	2.312	9.220	5.955	5.872	-1.550	9.657	3.558
8	13.128	5.039	9.618	5.854	7.047	-1.707	10.553	4.716
10	15.313	8.947	9.967	5.954	6.884	-2.129	11.873	5.475
12	15.663	13.118	9.989	6.265	4.982	-0.782	11.391	6.508
14	16.441	17.438	10.048	6.573	4.762	-0.218	11.546	7.398
16	18.233	20.720	10.134	6.900	4.359	0.871	12.220	8.259

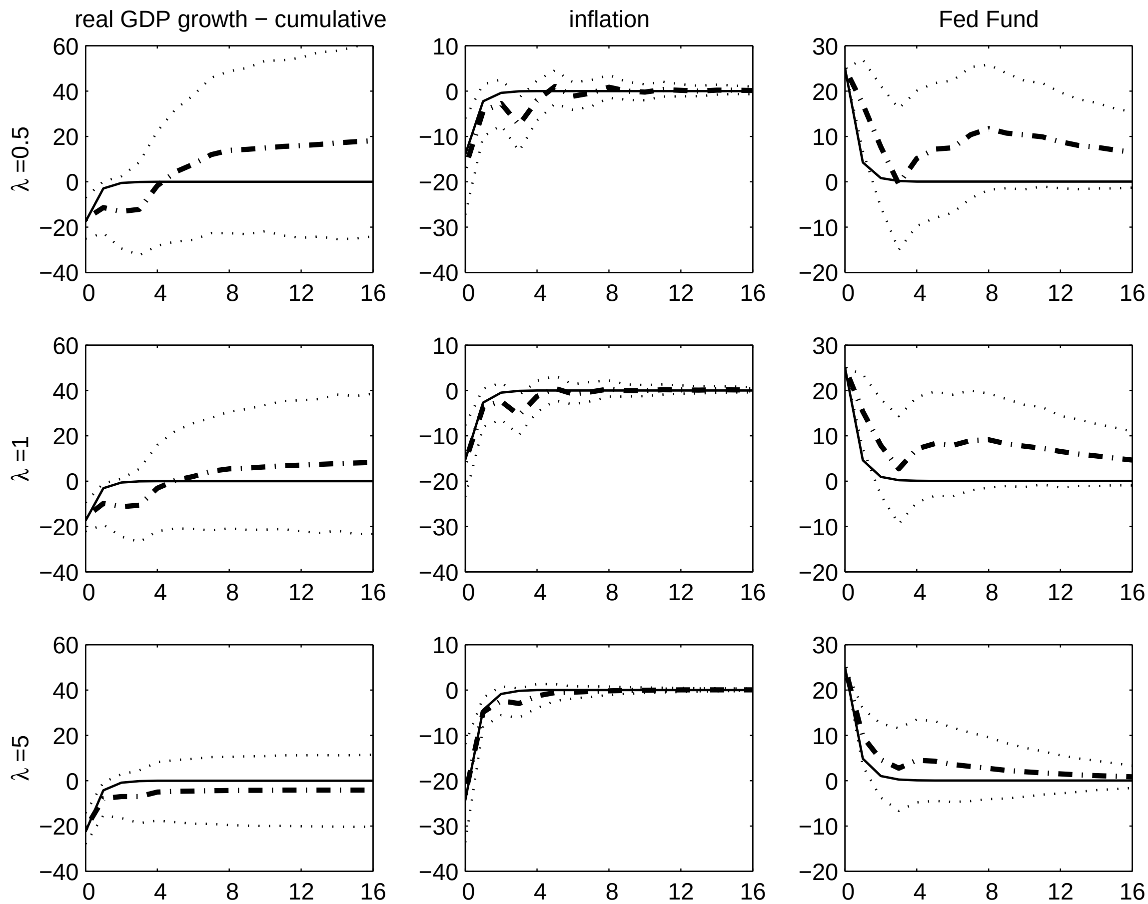
Notes: The rolling sample is 1975:III to 1997:III (90 periods). At each date in the sample, 80 observations are used in order to estimate the VAR. The forecasts are computed based an the values $\hat{\lambda}$ and $\hat{\iota}$ that have the highest posterior probability based on the estimation sample.

Figure 1: Forecasting performance as a function of the weight of the prior



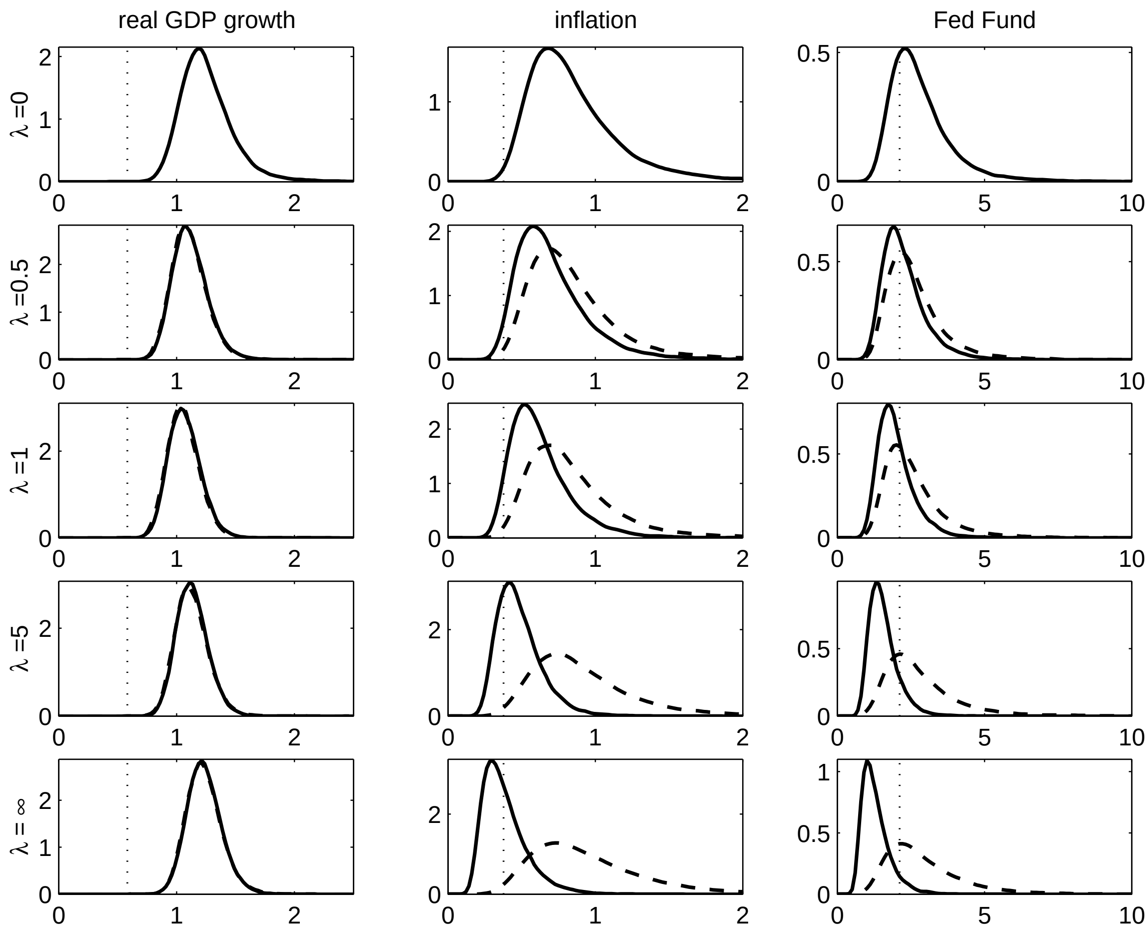
Notes: The plot shows the percentage gain (loss) in RMSEs relative to an unrestricted VAR. The rolling sample is 1975:III to 1997:III (90 periods). At each date in the sample, 80 observations are used in order to estimate the VAR.

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 $\psi_1 = 1.8$.