Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection School Of Economics

School of Economics

Bayesian Analysis of DSGE Models

Sungbae An
Singapore Management University, sungbae@smu.edu.sg

Frank Schorfheide

Follow this and additional works at: https://ink.library.smu.edu.sg/soe_research

Part of the Econometrics Commons

Bayesian Analysis of DSGE Models

Sungbae An, School of Economics and Social Sciences, Singapore Management University, Singapore

Frank Schorfheide, Department of Economics, University of Pennsylvania, Philadelphia, Pennsylvania, USA

Abstract

This paper reviews Bayesian methods that have been developed in recent years to estimate and evaluate dynamic stochastic general equilibrium (DSGE) models. We consider the estimation of linearized DSGE models, the evaluation of models based on Bayesian model checking, posterior odds comparisons, and comparisons to vector autoregressions, as well as the non-linear estimation based on a second-order accurate model solution. These methods are applied to data generated from correctly specified and misspecified linearized DSGE models and a DSGE model that was solved with a second-order perturbation method.

Keywords: Bayesian analysis, DSGE models, Model evaluation, Vector autoregressions

1. INTRODUCTION

Dynamic stochastic general equilibrium (DSGE) models are micro-founded optimization-based models that have become very popular in macroeconomics over the past 25 years. They are taught in virtually every Ph.D. program and represent a significant share of publications in macroeconomics. For a long time the quantitative evaluation of DSGE models was conducted without formal statistical methods. While DSGE models provide a complete multivariate stochastic process representation for the data, simple models impose very strong restrictions on actual time series and are in many cases rejected against less restrictive specifications such as vector autoregressions (VAR). Apparent model misspecifications were used as an argument in favor of informal calibration approaches along the lines of Kydland and Prescott (1982).

Subsequently, many authors have developed econometric frameworks that formalize aspects of the calibration approach by taking model misspecification explicitly into account. Examples are Smith (1993), Watson (1993), Canova (1994), DeJong et al. (1996), Diebold et

1

al. (1998), Geweke (1999b), Schorfheide (2000), Dridi et al. (2007), and Bierens (2007). At the same time, macroeconomists have improved the structural models and relaxed many of the misspecified restrictions of the first generation of DSGE models. As a consequence, more traditional econometric techniques have become applicable. The most recent vintage of DSGE models is not just attractive from a theoretical perspective but is also emerging as a useful tool for forecasting and quantitative policy analysis in macroeconomics. Moreover, owing to improved time series fit these models are gaining credibility in policy-making institutions such as central banks.

This paper reviews Bayesian estimation and evaluation techniques that have been developed in recent years for empirical work with DSGE models.[1] We focus on methods that are built around a likelihood function derived from the DSGE model. The econometric analysis has to cope with several challenges, including potential model misspecification and identification problems. We will illustrate how a Bayesian framework can address these challenges. Most of the techniques described in this article have been developed and applied in other papers. Our contribution is to give a unified perspective, by applying these methods successively to artificial data generated from a DSGE model and a VAR. We provide some evidence on the performance of Markov Chain Monte Carlo (MCMC) methods that have been applied to the Bayesian estimation of DSGE models. Moreover, we present new results on the use of first-order accurate versus second-order accurate solutions in the estimation of DSGE models.

The paper is structured as follows. Section 2 outlines two versions of a five-equation New Keynesian DSGE model along the lines of Woodford (2003) that differ with respect to the monetary policy rule. This model serves as the foundation for the current generation of large-scale models that are used for the analysis of monetary policy in academic and central bank circles. Section 3 discusses some preliminaries, in particular the challenges that have to be confronted by the econometric framework. We proceed by generating several data sets that are used subsequently for model estimation and evaluation. We simulate samples from the first-order accurate solution of the DSGE model presented in Section 2, from a modified version of the DSGE model in order to introduce misspecification, and from the second-order accurate solution of the benchmark DSGE model.

Owing to the computational burden associated with the likelihood evaluation for non-linear solutions of the DSGE model, most of the empirical literature has estimated linearized DSGE models. Section 4 describes a Random-Walk Metropolis (RWM) algorithm and an Importance Sampler (IS) algorithm that can be used to calculate the posterior moments of DSGE model parameters and transformations thereof. We compare the performance of the algorithms and provide an example in which they explore the posterior distribution only locally in the neighborhood of modes that are separated from each other by a deep valley in the surface of the posterior density.

Model evaluation is an important part of the empirical work with DSGE models. We consider three techniques in Section 5: posterior predictive model checking, model comparisons based on posterior odds, and comparisons of DSGE models to VARs. We illustrate these techniques by fitting correctly specified and misspecified DSGE models to artificial data. In Section 6 we construct posterior distributions for DSGE model parameters based on a second-order

---

[1] There is also an extensive literature on classical estimation and evaluation of DSGE models, but a detailed survey of these methods is beyond the scope of this article. The interested reader is referred to Kim and Pagan (1995) and the book by Canova (2007), which discusses both classical and Bayesian methods for the analysis of DSGE models.

accurate solution and compare them to posteriors obtained from a linearized DSGE model. Finally, Section 7 concludes and provides an outlook on future work.

2. A PROTOTYPICAL DSGE MODEL

Our model economy consists of a final goods producing firm, a continuum of intermediate goods producing firms, a representative household, and a monetary as well as a fiscal authority. This model has become a benchmark specification for the analysis of monetary policy and is analyzed in detail, for instance, in Woodford (2003). To keep the model specification simple, we abstract from wage rigidities and capital accumulation. More elaborate versions can be found in Smets and Wouters (2003).

2.1. The Agents and Their Decision Problems

The perfectly competitive, representative, final goods producing firm combines a continuum of intermediate goods indexed by $j \in [0, 1]$ using the technology

$$Y_t = \left(\int_0^1 Y_t(j)^{1-\nu} dj \right)^{\frac{1}{1-\nu}}. \tag{1}$$

Here $1/\nu > 1$ represents the elasticity of demand for each intermediate good. The firm takes input prices $P_t(j)$ and output prices P_t as given. Profit maximization implies that the demand for intermediate goods is

$$Y_t(j) = \left(\frac{P_t(j)}{P_t} \right)^{-1/\nu} Y_t. \tag{2}$$

The relationship between intermediate goods prices and the price of the final good is

$$P_t = \left(\int_0^1 P_t(j)^{\frac{\nu-1}{\nu}} dj \right)^{\frac{\nu}{\nu-1}}. \tag{3}$$

Intermediate good j is produced by a monopolist who has access to the linear production technology

$$Y_t(j) = A_t N_t(j), \tag{4}$$

where A_t is an exogenous productivity process that is common to all firms and $N_t(j)$ is the labor input of firm j. Labor is hired in a perfectly competitive factor market at the real wage W_t. Firms face nominal rigidities in terms of quadratic price adjustment costs

$$AC_t(j) = \frac{\phi}{2} \left(\frac{P_t(j)}{P_{t-1}(j)} - \pi \right)^2 Y_t(j), \tag{5}$$

where ϕ governs the price stickiness in the economy and π is the steady-state inflation rate associated with the final good. Firm j chooses its labor input $N_t(j)$ and the price $P_t(j)$ to maximize the present value of future profits

$$\mathbb{E}_t\left[\sum_{s=0}^{\infty}\beta^s Q_{t+s|t}\left(\frac{P_{t+s}(j)}{P_{t+s}}Y_{t+s}(j)-W_{t+s}N_{t+s}(j)-AC_{t+s}(j)\right)\right]. \qquad (6)$$

Here $Q_{t+s|t}$ is the time t value of a unit of the consumption good in period $t+s$ to the household, which is treated as exogenous by the firm.

The representative household derives utility from real money balances M_t/P_t and consumption C_t relative to a habit stock. We assume that the habit stock is given by the level of technology A_t. This assumption ensures that the economy evolves along a balanced growth path even if the utility function is additively separable in consumption, real money balances, and leisure. The household derives disutility from hours worked H_t and maximizes

$$\mathbb{E}_t\left[\sum_{s=0}^{\infty}\beta^s\left(\frac{(C_{t+s}/A_{t+s})^{1-\tau}-1}{1-\tau}+\chi_M\ln\left(\frac{M_{t+s}}{P_{t+s}}\right)-\chi_H H_{t+s}\right)\right], \qquad (7)$$

where β is the discount factor, $1/\tau$ is the intertemporal elasticity of substitution, and χ_M and χ_H are scale factors that determine steady-state real money balances and hours worked. We will set $\chi_H = 1$. The household supplies perfectly elastic labor services to the firms taking the real wage W_t as given. The household has access to a domestic bond market where nominal government bonds B_t are traded that pay (gross) interest R_t. Furthermore, it receives aggregate residual real profits D_t from the firms and has to pay lump-sum taxes T_t. Thus the household's budget constraint is of the form

$$P_t C_t + B_t + M_t - M_{t-1} + T_t = P_t W_t H_t + R_{t-1}B_{t-1} + P_t D_t + P_t SC_t, \qquad (8)$$

where SC_t is the net cash inflow from trading a full set of state-contingent securities. The usual transversality condition on asset accumulation applies, which rules out Ponzi schemes. Monetary policy is described by an interest rate feedback rule of the form

$$R_t = R_t^{*\,1-\rho_R}R_{t-1}^{\rho_R}e^{\varepsilon_{R,t}}, \qquad (9)$$

where $\varepsilon_{R,t}$ is a monetary policy shock and R_t^* is the (nominal) target rate. We consider two specifications for R_t^*, one in which the central bank reacts to inflation and deviations of output from potential output:

$$R_t^* = r\pi^*\left(\frac{\pi_t}{\pi^*}\right)^{\psi_1}\left(\frac{Y_t}{Y_t^*}\right)^{\psi_2} \qquad \text{(output gap rule specification)} \qquad (10)$$

and a second specification in which the central bank responds to deviations of output growth from its equilibrium steady-state γ:

$$R_t^* = r\pi^*\left(\frac{\pi_t}{\pi^*}\right)^{\psi_1}\left(\frac{Y_t}{\gamma Y_{t-1}}\right)^{\psi_2} \qquad \text{(output growth rule specification)}. \qquad (11)$$

Here r is the steady-state real interest rate, π_t is the gross inflation rate defined as $\pi_t = P_t/P_{t-1}$, and π^* is the target inflation rate, which in equilibrium coincides with the steady-state

inflation rate. Y_t* in (10) is the level of output that would prevail in the absence of nominal rigidities.

The fiscal authority consumes a fraction ζ_t of aggregate output Y_t, where $\zeta_t \in [0, 1]$ follows an exogenous process. The government levies a lump-sum tax (subsidy) to finance any shortfalls in government revenues (or to rebate any surplus). The government's budget constraint is given by

$$P_t G_t + R_{t-1} B_{t-1} = T_t + B_t + M_t - M_{t-1}, \tag{12}$$

where $G_t = \zeta_t Y_t$.

2.2 Exogenous Processes

The model economy is perturbed by three exogenous processes. Aggregate productivity evolves according to

$$\ln A_t = \ln \gamma - \ln A_{t-1} + \ln z_t, \qquad \text{where } \ln z_t = \rho_z \ln z_{t-1} + \epsilon_{z,t}. \tag{13}$$

Thus on average technology grows at the rate γ, and z_t captures exogenous fluctuations of the technology growth rate. Define $g_t = 1/(1 - \zeta_t)$. We assume that

$$\ln g_t = (1 - \rho_g)\ln g + \rho_g \ln g_{t-1} + \epsilon_{g,t}. \tag{14}$$

Finally, the monetary policy shock $\epsilon_{R,t}$ is assumed to be serially uncorrelated. The three innovations are independent of each other at all leads and lags and are normally distributed with means zero and standard deviations $\sigma_z, \sigma_g,$ and $\sigma_R,$ respectively.

2.3 Equilibrium Relationships

We consider the symmetric equilibrium in which all intermediate goods producing firms make identical choices so that the j subscript can be omitted. The market clearing conditions are given by

$$Y_t = C_t + G_t + AC_t \qquad \text{and} \qquad H_t = N_t. \tag{15}$$

Since the households have access to a full set of state-contingent claims, $Q_{t+s|t}$ in (6) is

$$Q_{t+s|t} = (C_{t+s}/C_t)^{-\tau} (A_t/A_{t+s})^{1-\tau}. \tag{16}$$

It can be shown that output, consumption, interest rates, and inflation have to satisfy the following optimality conditions

$$1 = \beta \mathbb{E}_t \left[\left(\frac{C_{t+1}/A_{t+1}}{C_t/A_t} \right)^{-\tau} \frac{A_t}{A_{t+1}} \frac{R_t}{\pi_{t+1}} \right] \tag{17}$$

$$1 = \frac{1}{v} \left[1 - \left(\frac{C_t}{A_t} \right)^{\tau} \right] + \phi(\pi_t - \pi) \left[\left(1 - \frac{1}{2v} \right) \pi_t + \frac{\pi}{2v} \right]$$

$$- \phi \beta \mathbb{E}_t \left[\left(\frac{C_{t+1}/A_{t+1}}{C_t/A_t} \right)^{-\tau} \frac{Y_{t+1}/A_{t+1}}{Y_t/A_t} (\pi_{t+1} - \pi) \pi_{t+1} \right]. \tag{18}$$

In the absence of nominal rigidities ($\phi = 0$) aggregate output is given by

$$Y_t^* = (1 - v)^{1/\tau} A_t g_t, \tag{19}$$

which is the target level of output that appears in the output gap rule specification. Since the non-stationary technology process A_t induces a stochastic trend in output and consumption, it is convenient to express the model in terms of detrended variables $c_t = C_t/A_t$ and $y_t = Y_t/A_t$. The model economy has a unique steady-state in terms of the detrended variables that is attained if the innovations $\epsilon_{R,t}$, $\epsilon_{g,t}$, and $\epsilon_{z,t}$ are zero at all times. The steady-state inflation π equals the target rate π_* and

$$r = \frac{\gamma}{\beta}, \qquad R = r\pi^*, \qquad c = (1 - v)^{1/\tau}, \qquad \text{and} \qquad y = g(1 - v)^{1/\tau}. \tag{20}$$

Let $\hat{x}_t = \ln(x_t/x)$ denote the percentage deviation of a variable x_t from its steady-state x. Then the model can be expressed as

$$1 = \beta \mathbb{E}_t [e^{-\tau \hat{c}_{t+1} + \tau \hat{c}_t + \hat{R}_t - \hat{z}_{t+1} - \hat{\pi}_{t+1}}] \tag{21}$$

$$\frac{1-v}{v\phi\pi^2} (e^{\tau \hat{c}_t} - 1) = (e^{\hat{\pi}_t} - 1) \left[\left(1 - \frac{1}{2v} \right) e^{\hat{\pi}_t} + \frac{1}{2v} \right]$$

$$- \beta \mathbb{E}_t \left[\left(e^{\hat{\pi}_{t+1}} - 1 \right) e^{-\tau \hat{c}_{t+1} + \tau \hat{c}_t + \hat{y}_{t+1} - \hat{y}_t + \hat{\pi}_{t+1}} \right] \tag{22}$$

$$e^{\hat{c}_t - \hat{y}_t} = e^{-\hat{g}_t} - \frac{\phi\pi^2 g}{2} (e^{\hat{\pi}_t} - 1)^2 \tag{23}$$

$$\widehat{R}_t = \rho_R \widehat{R}_{t-1} + (1 - \rho_R)\psi_1 \hat{\pi}_t + (1 - \rho_R)\psi_2(\hat{y}_t - \hat{g}_t) + \epsilon_{R,t} \tag{24}$$

$$\hat{g}_t = \rho_g \hat{g}_{t-1} + \epsilon_{g,t} \tag{25}$$

$$\hat{z}_t = \rho_z \hat{z}_{t-1} + \epsilon_{z,t}. \tag{26}$$

For the output growth rule specification, Equation (24) is replaced by

$$\widehat{R}_t = \rho_R \widehat{R}_{t-1} + (1 - \rho_R)\psi_1 \hat{\pi}_t + (1 - \rho_R)\psi_2 \left(\Delta \hat{y}_t + \hat{z}_t \right) + \epsilon_{R,t}. \tag{27}$$

2.4. Model Solutions

Equations (21) to (26) form a non-linear rational expectations system in the variables $\hat{y}_t$, $\hat{c}_t$, $\hat{\pi}_t$, $\widehat{R}_t$, $\hat{g}_t$, and $\hat{z}_t$ that is driven by the vector of innovations $\epsilon_t = [\epsilon_{R,t}, \epsilon_{g,t}, \epsilon_{z,t}]'$. This rational expectations system has to be solved before the DSGE model can be estimated. Define[2]

$$s_t = [\hat{y}_t, \hat{c}_t, \hat{\pi}_t, \widehat{R}_t, \epsilon_{R,t}, \hat{g}_t, \hat{z}_t]'.$$

The solution of the rational expectations system takes the form

$$s_t = \Phi(s_{t-1}, \epsilon_t; \theta). \tag{28}$$

From an econometric perspective, s_t can be viewed as a (partially latent) state vector in a non-linear state space model and (28) is the state transition equation.

A variety of numerical techniques are available to solve rational expectations systems. In the context of likelihood-based DSGE model estimation, linear approximation methods are very popular because they lead to a state–space representation of the DSGE model that can be analyzed with the Kalman filter. Linearization and straightforward manipulation of Equations (21) to (23) yields

$$\hat{y}_t = \mathbb{E}_t[\hat{y}_{t+1}] + \hat{g}_t - \mathbb{E}_t[\hat{g}_{t+1}] - \frac{1}{\tau}\left(\widehat{R}_t - \widehat{\mathbb{E}}_t[\pi_{t+1}] - \mathbb{E}_t[\hat{z}_{t+1}]\right) \tag{29}$$

$$\hat{\pi}_t = \beta\mathbb{E}_t[\hat{\pi}_{t+1}] + \kappa(\hat{y}_t - \hat{g}_t) \tag{30}$$

$$\hat{c}_t = \hat{y}_t - \hat{g}_t, \tag{31}$$

where

$$\kappa = \tau\frac{1-\nu}{\nu\pi^2\phi}. \tag{32}$$

Equations (29) to (31) combined with (24) to (26) and the trivial identity $\epsilon_{R,t} = \epsilon_{R,t}$ form a linear rational expectations system in s_t for which several solution algorithms are available, for instance, Blanchard and Kahn (1980), Binder and Pesaran (1997), King and Watson (1998), Uhlig (1999), Anderson (2000), Kim (2000), Christiano (2002), and Sims (2002). Depending on the parameterization of the DSGE model there are three possibilities: no stable rational expectations solution exists, the stable solution is unique (determinacy), or there are multiple stable solutions (indeterminacy). We will focus on the case of determinacy and restrict the parameter space accordingly. The resulting law of motion for the jth element of s_t takes the form

$$s_{j,t} = \sum_{i=1}^{J} \Phi_{j,i}^{(s)} s_{i,t-1} + \sum_{l=1}^{n} \Phi_{j,l}^{(\epsilon)} \epsilon_{l,t}, \qquad j = 1,\ldots,J. \tag{33}$$

---

[2] Under the output growth rule specification for R_t the vector s_t also contains $t-1$

Here J denotes the number of elements of the vector s_t and n is the number of shocks stacked in the vector ϵ_t. Here the coefficients $\Phi_{j,i}^{(s)}$ and $\Phi_{j,i}^{(\epsilon)}$ are functions of the structural parameters of the DSGE model.

While in many applications first-order approximations are sufficient, the higher-order refinement is an active area of research. For instance, if the goal of the analysis is to compare welfare across policies or market structures that do not have first-order effects on the model's steady-state or to study asset pricing implications of DSGE models, a more accurate solution may be necessary; see, for instance, Kim and Kim (2003) or Woodford, M. (2003). A simple example can illustrate this point. Consider a one-period model in which utility is a smooth function of consumption and is approximated as

$$U(\hat{c}) = U(0) + U^{(1)}(0)\hat{c} + \frac{1}{2}U^{(2)}(0)\hat{c}^2 + o(|\hat{c}|^2), \qquad (34)$$

where $\hat{c}$ is the percentage deviation of consumption from its steady-state and the superscript i denotes the ith derivative. Suppose that consumption is a smooth function of a zero mean random shock ϵ which is scaled by σ. The second-order approximation of $\hat{c}$ around the steady-state ($\sigma = 0$) is

$$\hat{c} = C^{(1)}(c)\sigma\epsilon + \frac{1}{2}C^{(2)}(c)\sigma^2\epsilon^2 + o_p(\sigma^2). \qquad (35)$$

If first-order approximations are used for both U and C, then the expected utility is simply approximated by the steady-state utility $U(0)$, which, for instance, in the DSGE model described above, is not affected by the coefficients ψ_1 and ψ_2 of the monetary policy rule (24). A second-order approximation of both U and C leads to

$$\mathbb{E}[U(\hat{c})] = U(0) + \frac{1}{2}U^{(2)}(0)C^{(1)^2}(c)\sigma^2 - \frac{1}{2}U^{(1)}(0)C^{(2)}(c)\sigma^2 + o(\sigma^2). \qquad (36)$$

Using a second-order expansion of the utility function together with a first-order expansion of consumption ignores the second term in (36) and is only appropriate if either $U^{(1)}(0)$ or $C^{(2)}(c)$ is zero. The first case arises if the marginal utility of consumption in the steady-state is zero, and the second case arises if the percentage deviation of consumption from the steady-state is a linear function of the shock.

A second-order accurate solution to the DSGE model can be obtained from a second-order expansion of the equilibrium conditions (21) to (26). Algorithms to construct such solutions have been developed by Judd (1998), Collard and Juillard (2001), Jin and Judd (2002), Schmitt-Grohé and Uribe (2006), Kim et al. (2005), and Swanson et al. (2005). The resulting state transition equation can be expressed as

$$s_{j,t} = \Phi_j^{(0)} + \sum_{i=1}^{J} \Phi_{j,it}^{(s)} s_{i,t-1} + \sum_{l=1}^{n} \Phi_{j,it}^{(\epsilon)} \epsilon_{l,t} + \sum_{i=1}^{J} \sum_{l=1}^{J} \Phi_{j,il}^{(ss)} s_{i,t-1} s_{l,t-1}$$

$$+ \sum_{i=1}^{J} \sum_{l=1}^{n} \Phi_{j,il}^{(s\epsilon)} s_{i,t-1} \epsilon_{l,t} + \sum_{i=1}^{n} \sum_{l=1}^{n} \Phi_{j,il}^{(\epsilon\epsilon)} \epsilon_{i,t} \epsilon_{l,t}. \tag{37}$$

As before, the coefficients are functions of the parameters of the DSGE model. For the subsequent analysis we use Sims' (2002) procedure to compute a first-order accurate solution of the DSGE model and Schmitt-Grohé and Uribe's (2006) algorithm to obtain a second-order accurate solution.

While perturbation methods approximate policy functions only locally, there are several global approximation schemes, including projection methods such as the finite-elements method and Chebyshev–polynomial method on the spectral domain. Judd (1998) covers various solution methods, and Taylor and Uhlig (1990), Den Haan and Marcet (1994), and Aruoba et al. (2004) compare the accuracy of alternative solution methods.

2.5. Measurement Equations

The model is completed by defining a set of measurement equations that relate the elements of s_t to a set of observables. We assume that the time period t in the model corresponds to one quarter and that the following observations are available for estimation: quarter-to-quarter per capita GDP growth rates (YGR), annualized quarter-to-quarter inflation rates (INFL), and annualized nominal interest rates (INT). The three series are measured in percentages, and their relationship to the model variables is given by the set of equations

$$\begin{aligned} YGR_t &= \gamma^{(Q)} + 100(\hat{y}_t - \hat{y}_{t-1} + \hat{z}_t) \\ INFL_t &= \pi^{(A)} + 400\hat{\pi}_t \\ INT_t &= \pi^{(A)} + r^{(A)} + 4\gamma^{(Q)} + 400\widehat{R}_t. \end{aligned} \tag{38}$$

The parameters $\gamma^{(Q)}$, $\pi^{(A)}$, and $r^{(A)}$ are related to the steady states of the model economy as

$$\gamma = 1 + \frac{\gamma^{(Q)}}{100}, \qquad \beta = \frac{1}{1 + r^{(A)}/100}, \qquad \pi = 1 + \frac{\pi^{(A)}}{100}.$$

The structural parameters are collected in the vector θ. Since in the first-order approximation the parameters ν and ϕ are not separately identifiable, we express the model in terms of κ, defined in (32). Let

$$\theta = [\tau, \kappa, \psi_1, \psi_2, \rho_R, \rho_g, \rho_z, r^{(A)}, \pi^{(A)}, \gamma^{(Q)}, \sigma_R, \sigma_g, \sigma_z]'.$$

For the quadratic approximation the composite parameter κ will be replaced by either (ψ, v) or alternatively by (κ, v). Moreover, θ will be augmented with the steady-state ratio c/y, which is equal to $1/g$.

3. PRELIMINARIES

Numerous formal and informal econometric procedures have been proposed to parameterize and evaluate DSGE models, ranging from calibration, e.g., Kydland and Prescott (1982) over generalized method of moments (GMM) estimation of equilibrium relationships, e.g., Christiano and Eichenbaum (1992), minimum distance estimation based on the discrepancy among VAR and DSGE model impulse response functions, e.g., Rotemberg and Woodford (1997), to full-information likelihood-based estimation as in Altug (1989), McGrattan (1994), Leeper and Sims (1994), and Kim (2000). Much of the methodological debate surrounding the various estimation (and model evaluation) techniques is summarized in papers by Kydland and Prescott (1996), Hansen and Heckman (1996), and Sims (1996).

We focus on Bayesian estimation of DSGE models, which has three main characteristics. First, unlike GMM estimation based on equilibrium relationships such as the consumption Euler equation (21), the price setting equation of the intermediate goods producing firms (22), or the monetary policy rule (24), the Bayesian analysis is system-based and fits the solved DSGE model to a vector of aggregate time series. Second, the estimation is based on the likelihood function generated by the DSGE model rather than, for instance, the discrepancy between DSGE model responses and VAR impulse responses. Third, prior distributions can be used to incorporate additional information into the parameter estimation. Any estimation and evaluation method is confronted with the following challenges: potential model misspecification and possible lack of identification of parameters of interest. We will subsequently elaborate these challenges and discuss how a Bayesian approach can be used to cope with them.

Throughout the paper we will use the following notation: the $n \times 1$ vector y_t stacks the time t observations that are used to estimate the DSGE model. In the context of the model developed in Section 2 y_t is composed of YGR_t, $INFL_t$, INT_t. The sample ranges from $t = 1$ to T and the sample observations are collected in the matrix Y with rows y_t'. We denote the prior density by $p(\theta)$, the likelihood function by $\mathcal{L}(\theta \mid Y)$, and the posterior density by $p(\theta \mid Y)$.

3.1. Potential Model Misspecification

If one predicts a vector of time series y_t, for instance composed of output growth, inflation, and nominal interest rates, by a function of past y_t's, then the resulting forecast error covariance matrix is non-singular. Hence any DSGE model that generates a rank-deficient covariance matrix for y_t is clearly at odds with the data and suffers from an obvious form of misspecification. This singularity is an obstacle to likelihood estimation. Hence one branch of the literature has emphasized procedures that can be applied despite the singularity, whereas

the other branch has modified the model specification to remove the singularity by adding so-called measurement errors, e.g., Sargent (1989), Altug (1989), Ireland (2004), or additional structural shocks as in Leeper and Sims (1994) and more recently Smets and Wouters (2003). In this paper we pursue the latter approach by considering a model in which the number of structural shocks (monetary policy shock, government spending shock, technology growth shock) equals the number of observables (output growth, inflation, interest rates) to which the model is fitted.

A second source of misspecification that is more difficult to correct is potentially invalid cross-coefficient restrictions on the time series representation of y_t generated by the DSGE model. Invalid restrictions manifest themselves in poor out-of-sample fit relative to more densely parameterized reference models such as VARs. Del Negro et al. (2006) document that even an elaborate DSGE model with capital accumulation and various nominal and real frictions has difficulties attaining the fit achieved with VARs that have been estimated with well-designed shrinkage methods. While the primary research objectives in the DSGE model literature is to overcome discrepancies between models and reality, it is important to have empirical strategies available that are able to cope with potential model misspecification.

Once one acknowledges that the DSGE model provides merely an approximation to the law of motion of the time series y_t, then it seems reasonable to assume that there need not exist a single parameter vector θ_0 that delivers, say, the "true" intertemporal substitution elasticity or price adjustment costs and, simultaneously, the most precise impulse responses to a technology or monetary policy shock. Each estimation method is associated with a particular measure of discrepancy between the 'true' law of motion and the class of approximating models. Likelihood-based estimators, for instance, asymptotically minimize the Kullback–Leibler distance (see White, 1982).

One reason that pure maximum likelihood estimation of DSGE models has not turned into the estimation method of choice is the "dilemma of absurd parameter estimates." Estimates of structural parameters generated with maximum likelihood procedures based on a set of observations Y are often at odds with additional information that the research may have. For example, estimates of the discount factor β should be consistent with our knowledge about the average magnitude of real interest rates, even if observations on interest rates are not included in the estimation sample Y. Time series estimates of aggregate labor supply elasticities or price adjustment costs should be broadly consistent with microlevel evidence. However, due to the stylized nature and the resulting misspecification of most DSGE models, the likelihood function often peaks in regions of the parameter space that appear to be inconsistent with extraneous information.

In a Bayesian framework, the likelihood function is reweighted by a prior density. The prior can bring to bear information that is not contained in the estimation sample Y. Since priors are always subject to revision, the shift from prior to posterior distribution can be an indicator of the tension between different sources of information. If the likelihood function peaks at a value that is at odds with the information that has been used to construct the prior distribution, then the marginal data density of the DSGE model, defined as

$$p(Y) = \int \mathcal{L}(\theta \mid Y) p(\theta) d\theta, \tag{39}$$

will be low compared to, say, a VAR, and in a posterior odds comparison the DSGE model will automatically be penalized for not being able to reconcile the two sources of information with a single set of parameters. Section 5 will discuss several methods that have been proposed to assess the fit of DSGE models: posterior odds comparisons of competing model specifications, posterior predictive checks, and comparisons to reference models that relax some of the cross-coefficient restrictions generated by the DSGE models.

3.2. Identification

Identification problems can arise owing to a lack of informative observations or, more fundamentally, from a probability model that implies that different values of structural parameters lead to the same joint distribution for the observables *Y*. At first glance the identification of DSGE model parameters does not appear to be problematic. The parameter vector θ is typically low dimensional compared to VARs and the model imposes tight restrictions on the time series representation of *Y*. However, recent experience with the estimation of New Keynesian DSGE models has cast some doubt on this notion and triggered a more careful assessment. Lack of identification is documented in papers by Beyer and Farmer (2004) and Canova and Sala (2005). The former paper provides an algorithm to construct families of observationally equivalent linear rational expectations models, whereas the latter paper compares the informativeness of different estimators with respect to key structural parameters in a variety of DSGE models.

The delicate identification problems that arise in rational expectations models can be illustrated in a simple example adopted from Lubik and Schorfheide (2006). Consider the following two models, in which y_t is the observed endogenous variable and u_t is an unobserved shock process. In model $\mathcal{M}_1$, the u_t's are serially correlated:

$$\mathcal{M}_1 : y_t - \frac{1}{\alpha} \mathbb{E}_t[y_{t+1}] + u_t, \qquad u_t - \rho u_{t-1} + \epsilon_t, \qquad \epsilon_t \sim iid \left(0, (1 - \rho/\alpha)^2\right). \tag{40}$$

In model $\mathcal{M}_2$ the shocks are serially uncorrelated, but we introduce a backward-looking term ϕy_{t-1} on the right-hand side to generate persistence:

$$\mathcal{M}_2 : y_t = \frac{1}{\alpha} \mathbb{E}_t[y_{t+1}] + \phi y_{t-1} + u_t, \qquad u_t = \epsilon_t,$$

$$\epsilon_t \sim iid \left(0, \left[\frac{\alpha + \sqrt{\alpha^2 - 4\phi\alpha}}{2\alpha}\right]^2\right). \tag{41}$$

For both specifications, the law of motion of y_t is

$$y_t = \psi y_{t-1} + \eta_t, \qquad \eta_t \sim iid(0,1). \tag{42}$$

Under restrictions on the parameter spaces that guarantee uniqueness of a stable rational expectations solution[3] we obtain the following relationships between ψ and the structural parameters:

$$\mathcal{M}_1 : \psi = \rho, \qquad \mathcal{M}_2 : \psi = \frac{1}{2}(\alpha - \sqrt{\alpha^2 - 4\phi\alpha}).$$

In model $\mathcal{M}_1$ the parameter α is not identifiable, and in model $\mathcal{M}_2$, the parameters α and ϕ are not separately identifiable. Moreover, models $\mathcal{M}_1$ and $\mathcal{M}_2$ are observationally equivalent. The likelihood functions of $\mathcal{M}_1$ and $\mathcal{M}_2$ have ridges that can cause serious problems for numerical optimization procedures. The calculation of a valid confidence set is challenging since it is difficult to trace out the parameter subspace along which the likelihood function is constant.

While Bayesian inference is based on the likelihood function, even a weakly informative prior can introduce curvature into the posterior density surface that facilitates numerical maximization and the use of MCMC methods. Consider model $\mathcal{M}_1$. Here $\theta = [\rho, \alpha]'$ and the likelihood function can be written as $\mathcal{L}(\rho, \alpha \,|\, Y) = \overline{\mathcal{L}}(\rho)$. Straightforward manipulations of the Bayes theorem yield

$$p(\rho, \alpha \,|\, Y) = p(\rho \,|\, Y)p(\alpha \,|\, \rho). \tag{43}$$

Thus the prior distribution is not updated in directions of the parameter space in which the likelihood function is flat. This is of course well known in Bayesian econometrics; see Poirier (1998) for a discussion.

It is difficult to detect directly identification problems in large DSGE models, since the mapping from the vector of structural parameters θ into the state–space representation that determines the joint probability distribution of Y is highly non-linear and typically can only be evaluated numerically. The posterior distribution is well defined as long as the joint prior distribution of the parameters is proper. However, lack of identification provides a challenge for scientific reporting, as the audience typically would like to know what features of the posterior are generated by the prior rather than the likelihood. A direct comparison of priors and posteriors can often provide valuable insights about the extent to which data provide information about the parameters of interest.

---

[3] To ensure determinacy we impose $\alpha > 1$ in $\mathcal{M}_1$ and $|\alpha - \sqrt{\alpha^2 - 4\phi\alpha}| < 2$ and $|\alpha + \sqrt{\alpha^2 - 4\phi\alpha}| > 2$ in $\mathcal{M}_2$.

3.3. Priors

As indicated in our previous discussion, prior distributions will play an important role in the estimation of DSGE models. They might downweigh regions of the parameter space that are at odds with observations not contained in the estimation sample Y. They might also add curvature to a likelihood function that is (nearly) flat in some dimensions of the parameter space and therefore strongly influence the shape of the posterior distribution.[4] While, in principle, priors can be gleaned from personal introspection to reflect strongly held beliefs about the validity of economic theories, in practice most priors are chosen based on some observations.

For instance, Lubik and Schorfheide (2006) estimate a two-country version of the model described in Section 2. Priors for the autocorrelations and standard deviations of the exogenous processes, the steady-state parameters $\gamma^{(Q)}$, $\pi^{(A)}$, and $r^{(A)}$, as well as the standard deviation of the monetary policy shock, are quantified based on regressions run on pre-(estimation)-sample observations of output growth, inflation, and nominal interest rates. The priors for the coefficients in the monetary policy rule are loosely centered around values typically associated with the Taylor rule. The prior for the parameter that governs price stickiness is chosen based on microevidence on price setting behavior provided, for instance, in Bils and Klenow (2004)**.** To fine-tune the prior distribution, in particular the distribution of shock standard deviations, it is often helpful to simulate the prior predictive distribution for various sample moments and check that the prior does not place little or no mass in a neighborhood of important features of the data. Such an occurrence would suggest that the model is incapable of explaining salient data properties. A formalization of such a prior predictive check can be found in Geweke (2005).

Table 2 lists the marginal prior distributions for the structural parameters of the DSGE model, that we will use in the subsequent analysis. These priors are adopted from Lubik and Schorfheide (2006). For convenience, it is assumed that all parameters are a priori independent. In applications in which the independence assumption is unreasonable one could derive parameter transformations, such as steady rate ratios, autocorrelations, or relative volatilities, and specify independent priors on the transformed parameters, which induce dependent priors for the original parameters. As mentioned before, rational expectations models can have multiple equilibria. While this may be of independent interest we do not pursue this direction in this paper. Hence, the prior distribution is truncated at the boundary of the determinacy region. Prior to the truncation the distribution specified in Table 2 places about 2% of its mass on parameter values that imply indeterminacy. The parameters ν (conditional on κ) and $1/g$ only affect the second-order approximation of the DSGE model.

---

[4] The role of priors in DSGE model estimation is markedly different from the role of priors in VAR estimation. In the latter case priors are essentially used to reduce the dimensionality of the econometric model and hence the sampling variability of the parameter estimates.

3.4. The Road Ahead

Throughout this paper we are estimating and evaluating versions of the DSGE model presented in Section 2 based on simulated data. Table 1 provides a characterization of the model specifications and data sets.

Model $\mathcal{M}_1$ is the benchmark specification of the DSGE model in which monetary policy follows an interest-rate rule that reacts to the output gap. $\mathcal{M}_1(L)$ is approximated by a linear rational expectations system, whereas $\mathcal{M}_1(Q)$ is solved with a second-order perturbation method. In $\mathcal{M}_2(L)$ we replace the output gap in the interest-rate feedback rule by output growth. $\mathcal{M}_3$ and $\mathcal{M}_4$ are identical to $\mathcal{M}_1(L)$, except that in one case we impose that the prices are nearly flexible ($\kappa = 5$) and in the other case the central bank does not respond to output ($\psi_2 = 0$). In $\mathcal{M}_5(L)$ we replace the log-linearized consumption Euler equation by an equation that includes lagged output on the right-hand side. Loosely speaking, this specification can be motivated with a model in which agents derive utility from the difference between individual consumption and the overall level of consumption in the previous period.

TABLE 1 Model specifications and data sets

$\mathcal{M}_1(L)$	Benchmark Model with output gap rule, consists of Eqs. (21)–(26), solved by first-order approximation.
$\mathcal{M}_1(Q)$	Benchmark Model with output gap rule, consists of Eqs. (21)–(26), solved by second-order approximation.
$\mathcal{M}_2(L)$	DSGE Model with output growth rule, consists of Eqs. (21)–(26), however, Eq. (24) is replaced by Eq. (27), solved by first-order approximation.
$\mathcal{M}_3(L)$	Same as $\mathcal{M}_1(L)$, except that prices are nearly flexible: $\kappa = 5$.
$\mathcal{M}_4(L)$	Same as $\mathcal{M}_1(L)$, except that central bank does not respond to the output gap: $\psi_2 = 0$.
$\mathcal{M}_5(L)$	Same as $\mathcal{M}_1(L)$, except in first-order approximation Eq. (29)is replaced by $(1+h)\hat{y}_t = \hat{\mathbb{E}}_t[y_{t+1}] + h\hat{y}_{t-1} + \hat{g}_t - \mathbb{E}_t[\hat{g}_{t+1}] - \frac{1}{\tau}(\hat{R}_t - \hat{\mathbb{E}}_t[\pi_{t+1}] - \mathbb{E}[\hat{\varepsilon}_{t+1}])$ with $h = .95$.
$\mathcal{D}_1(L)$	80 observations generated with $\mathcal{M}_1(L)$
$\mathcal{D}_1(Q)$	80 observations generated with $\mathcal{M}_1(Q)$
$\mathcal{D}_2(L)$	80 observations generated with $\mathcal{M}_2(L)$
$\mathcal{D}_5(L)$	80 observations generated with $\mathcal{M}_5(L)$

Conditional on parameter vectors reported in Tables 2 and 3 we generate four data sets with 80 observations each. We use the labels $\mathcal{D}_1(L)$, $\mathcal{D}_1(Q)$, $\mathcal{D}_2(L)$, and $\mathcal{D}_5(L)$ to denote data sets generated from $\mathcal{M}_1(L)$, $\mathcal{M}_1(Q)$, $\mathcal{M}_2(L)$, and $\mathcal{M}_5(L)$, respectively. By and large, the values for the DSGE model parameters resemble empirical estimates obtained from post-1982 U.S. data. The sample size is fairly realistic for the estimation of monetary DSGE model. Many industrialized countries experienced a high inflation episode in the 1970s that ended with a disinflation in the 1980s. Subsequently, the level and the variability of inflation and interest rates have been fairly stable, so that it is not uncommon to estimate constant coefficient DSGE models based on data sets that begin in the early 1980s.

TABLE 2 Prior distribution and DGPs—linear analysis (Sections 4 and 5)

Name	Domain	Density	Prior Para (1)	Prior Para (2)	DGP $\mathcal{D}_1(L), \mathcal{D}_2(L), \mathcal{D}_5(L)$
τ	$\mathbb{R}^+$	Gamma	2.00	.50	2.00
κ	$\mathbb{R}^+$	Gamma	.20	.10	.15
ψ_1	$\mathbb{R}^+$	Gamma	1.50	.25	1.50
ψ_2	$\mathbb{R}^+$	Gamma	.50	.25	1.00
ρ_R	[0, 1)	Beta	.50	.20	.60
ρ_G	[0, 1)	Beta	.80	.10	.95
ρ_Z	[0, 1)	Beta	.66	.15	.65
$r^{(A)}$	$\mathbb{R}^+$	Gamma	.50	.50	.40
$\pi^{(A)}$	$\mathbb{R}^+$	Gamma	7.00	2.00	4.00
$\gamma^{(Q)}$	$\mathbb{R}$	Normal	.40	.20	.50
$100\,\sigma_R$	$\mathbb{R}^+$	InvGamma	.40	4.00	.20
$100\,\sigma_G$	$\mathbb{R}^+$	InvGamma	1.00	4.00	.80
$100\,\sigma_Z$	$\mathbb{R}^+$	InvGamma	.50	4.00	.45

Notes: Paras (1) and (2) list the means and the standard deviations for Beta, Gamma, and Normal distributions; the upper and lower bound of the support for the Uniform distribution; *s* and ν for the Inverse Gamma distribution, where $p\,\mathcal{I}_G(\sigma \mid \nu, s) \propto \sigma^{-\nu-1} e^{-\nu s^2/2\sigma^2}$. The effective prior is truncated at the boundary of the determinacy region.

TABLE 3 Prior distribution and DGP—Nonlinear analysis (Section 6)

Name	Domain	Density	Prior Para (1)	Prior Para (2)	DGP $\mathcal{D}_1(Q)$
τ	$\mathbb{R}^+$	Gamma	2.00	.50	2.00
κ	$\mathbb{R}^+$	Gamma	.30	.20	.33
ψ_1	$\mathbb{R}^+$	Gamma	1.50	.25	1.50
ψ_2	$\mathbb{R}^+$	Gamma	.50	.25	.125
ρ_R	[0, 1)	Beta	.50	.20	.75
ρ_G	[0, 1)	Beta	.80	.10	.95
ρ_Z	[0, 1)	Beta	.66	.15	.90
$r^{(A)}$	$\mathbb{R}^+$	Gamma	.80	.50	1.00
$\pi^{(A)}$	$\mathbb{R}^+$	Gamma	4.00	2.00	3.20
$\gamma^{(Q)}$	$\mathbb{R}$	Normal	.40	.20	.55
$100\sigma_R$	$\mathbb{R}^+$	InvGamma	.30	4.00	.20
$100\,\sigma_G$	$\mathbb{R}^+$	InvGamma	.40	4.00	.60
$100\,\sigma_Z$	$\mathbb{R}^+$	InvGamma	.40	4.00	.30
ν	[0, 1)	Beta	.10	.05	.10
$1/g$	[0, 1)	Beta	.85	.10	.85

Notes: See Table 2. Parameters ν and 1/*g* only affect the second-order accurate solution of the DSGE model.

Section 4 illustrates the estimation of linearized DSGE models under correct specification, that is, we construct posteriors for $\mathcal{M}_1(L)$ and $\mathcal{M}_2(L)$ based on the data sets $\mathcal{D}_1(L)$ and $\mathcal{D}_2(L)$. Section 5 considers model evaluation techniques. We begin with posterior predictive checks in Section 5.1, which are implemented based on $\mathcal{M}_1(L)$ for data sets $\mathcal{D}_1(L)$ (correct specification of DSGE model) and $\mathcal{D}_5(L)$ (misspecification of the consumption Euler equation). In Section 5.2 we proceed with the calculation of posterior model probabilities for specifications $\mathcal{M}_1(L)$, $\mathcal{M}_3(L)$, and $\mathcal{M}_4(L)$ conditional on the data set $\mathcal{D}_1(L)$. Subsequently in Section 5.3 we compare $\mathcal{M}_1(L)$ to a vector autoregressive specification using $\mathcal{D}_1(L)$ (correct specification) and $\mathcal{D}_5(L)$ (misspecification). Finally, in Section 6 we study the estimation of DSGE models solved with a second-order perturbation method. We compare posteriors for $\mathcal{M}_1(Q)$ and $\mathcal{M}_1(L)$ conditional on $\mathcal{D}_1(Q)$.

4. ESTIMATION OF LINEARIZED DSGE MODELS

We will begin by describing two algorithms that can be used to generate draws from the posterior distribution of θ and subsequently illustrate their performance in the context of models $\mathcal{M}_1(L)/\mathcal{D}_1(L)$ and $\mathcal{M}_2(L)/\mathcal{D}_2(L)$.

4.1. Posterior Computations

We consider an RWM algorithm and an IS algorithm to generate draws from the posterior distribution of θ. Both algorithms require the evaluation of $\mathcal{L}(\theta \mid Y)p(\theta)$. The computation of the non-normalized posterior density proceeds in two steps. First, the linear rational expectations system is solved to obtain the state transition equation (33). If the parameter value θ implies indeterminacy (or non-existence of a stable rational expectations solution), then $\mathcal{L}(\theta \mid Y)p(\theta)$ is set to zero. If a unique stable solution exists, then the Kalman filter[5] is used to evaluate the likelihood function associated with the linear state–space system (33) and (38). Since the prior is generated from well-known densities, the computation of $p(\theta)$ is straightforward.

The RWM algorithm belongs to the more general class of Metropolis–Hastings algorithms. This class is composed of universal algorithms that generate Markov chains with stationary distributions that correspond to the posterior distributions of interest. A first version of such an algorithm had been constructed by Metropolis et al. (1953) to solve a minimization problem and was later generalized by Hastings (1970). Chib and Greenberg (1995) provide an excellent introduction to Metropolis–Hastings algorithms. The RWM algorithm was first

---

[5] Since according to our model s_t is stationary, the Kalman filter can be initialized with the unconditional distribution of s_t. To make the estimation in Section 4 comparable to the DSGE-VAR analysis presented in Section 5.3 we adjust the likelihood function to condition on the first four observations in the sample: $\mathcal{L}(\theta \mid Y)/\mathcal{L}(\theta \mid y_1,\ldots,y_4)$. These four observations are later used to initialize the lags of a VAR.

used to generate draws from the posterior distribution of DSGE model parameters by Schorfheide (2000).[6]

Random-Walk Metropolis (RWM) Algorithm

1. Use a numerical optimization routine to maximize ln $\mathscr{L}(\theta \mid Y) + \ln p(\theta)$. Denote the posterior mode by $\tilde{\theta}$.

2. Let $\tilde{\Sigma}$ be the inverse of the Hessian computed at the posterior mode $\tilde{\theta}$.

3. Draw $\theta^{(0)}$ from $\mathscr{N}(\tilde{\theta}, c_0^2 \tilde{\Sigma})$ or directly specify a starting value.

4. For $s = 1,\ldots,n_{sim,}$ draw ϑ from the proposal distribution $\mathscr{N}(\theta^{(s-1)}, c^2 \tilde{\Sigma})$. The jump from $\theta^{(s-1)}$ is accepted ($\theta^{(s)} = \vartheta$) with probability min$\{1, r(\theta^{(s-1)}, \vartheta \mid Y)\}$ and rejected ($\theta^{(s)} = \theta^{(s-1)}$) otherwise. Here

$$r(\theta^{(s-1)}, \vartheta \mid Y) - \frac{\mathscr{L}(\vartheta \mid Y) p(\vartheta)}{\mathscr{L}(\theta^{(s-1)} \mid Y) p(\theta^{(s-1)})}.$$

5. Approximate the posterior expected value of a function $h(\theta)$ by $\frac{1}{n_{sim}} \sum_{s=1}^{n_{sim}} h(\theta^{(s)})$.

Under fairly general regularity conditions, e.g., Walker (1969), Crowder (1988), and Kim (1998), the posterior distribution of θ will be asymptotically normal. The algorithm constructs a Gaussian approximation around the posterior mode and uses a scaled version of the asymptotic covariance matrix as the covariance matrix for the proposal distribution. This allows for an efficient exploration of the posterior distribution at least in the neighborhood of the mode.

The maximization of the posterior density kernel is carried out with a version of the BFGS quasi-Newton algorithm, written by Chris Sims for the maximum likelihood estimation of a DSGE model conducted in Leeper and Sims (1994). The algorithm uses a fairly simple line search and randomly perturbs the search direction if it reaches a cliff caused by nonexistence or non-uniqueness of a stable rational expectations solution for the DSGE model. Prior to the numerical maximization we transform all parameters so that their domain is unconstrained. This parameter transformation is only used in Step 1 of the RWM algorithm. The elements of the Hessian matrix in Step 2 are computed numerically for various values of the increment $d\theta$. There typically exists a range for $d\theta$ over which the derivatives are stable. As $d\theta$ approaches zero the numerical derivatives will eventually deteriorate owing to inaccuracies in

---

[6] This version of the algorithm is available in the user-friendly DYNARE (2005) package, which automates the Bayesian estimation of linearized DSGE models and provides routines to solve DSGE models with higher-order perturbation methods.

the evaluation of the objective function. While Steps 1 and 2 are not necessary for the implementation of the RWM algorithm, they are often helpful.

The RWM algorithm generates a sequence of dependent draws from the posterior distribution of θ that can be averaged to approximate posterior moments. Geweke (1999a, 2005) reviews regularity conditions that guarantee the convergence of the Markov chain generated by Metropolis–Hastings algorithms to the posterior distribution of interest and the convergence of $\frac{1}{n_{sim}} \sum_{s=1}^{n_{sim}} h(\theta^{(s)})$ to the posterior expectations $\mathbb{E}[h(\theta) \mid Y]$.

DeJong et al. (2000) used an IS algorithm to calculate posterior moments of the parameters of a linearized stochastic growth model. The idea of the algorithm is based on the identity

$$\mathbb{E}[h(\theta) \mid Y] = \int h(\theta) p(\theta \mid Y) d\theta = \int \frac{h(\theta) p(\theta \mid Y)}{q(\theta)} q(\theta) d\theta.$$

Draws from the posterior density $p(\theta \mid Y)$ are replaced by draws from the density $q(\theta)$ and reweighted by the importance ratio $p(\theta \mid Y)/q(\theta)$ to obtain a numerical approximation of the posterior moment of interest. Hammersley and Handscomb (1964) were among the first to propose this method and Geweke (1989) provides important convergence results. The particular version of the IS algorithm used subsequently is of the following form.

Importance Sampling (IS) Algorithm

1. Use a numerical optimization routine to maximize $\ln \mathscr{L}(\theta \mid Y) + \ln p(\theta)$. Denote the posterior mode by .

2. Let $\tilde{\Sigma}$ be the inverse of the Hessian computed at the posterior mode $\square$.

3. Let $q(\theta)$ be the density of a multivariate t-distribution with mean $\square$, scale matrix $c^2 \tilde{\Sigma}$, and ν degrees of freedom.

4. For $s = 1,\ldots,n_{sim}$ generate draws $\theta^{(s)}$ from $q(\theta)$.

5. Compute $\bar{w}_s = \mathscr{L}(\theta^{(s)} \mid Y) p(\theta^{(s)})/q(\theta^{(s)})$ and $w_s = \tilde{w}_s / \sum_{s=1}^{n_{sim}} \tilde{w}_s$.

6. Approximate the posterior expected value of a function $h(\theta)$ by $\sum_{s=1}^{n_{sim}} w(\theta^{(s)}) h(\theta^{(s)})$.

The accuracy of the IS approximation depends on the similarity between posterior kernel and importance density. If the two are equal, then the importance weights w_s are constant and one averages independent draws to approximate the posterior expectations of interest. If the importance density concentrates its mass in a region of the parameter space in which the

posterior density is very low, then most of the w_s's will be much smaller than $1/n_{sim}$ and the approximation method is inefficient. We construct a Gaussian approximation of the posterior near the mode, scale the asymptotic covariance matrix, and replace the normal distribution with a fat-tailed t-distribution to implement the algorithm.

4.2. Simulation Results for $\mathcal{M}_1(L)/\mathcal{D}_1(L)$

We begin by estimating the log-linearized output gap rule specification $\mathcal{M}_1(L)$ based on data set $\mathcal{D}_1(L)$. We use the RWM algorithm ($c_0 = 1$, $c = 0.3$) to generate 1 million draws from the posterior distribution. The rejection rate is about 45%. Moreover, we generate 200 independent draws from the truncated prior distribution. Figure 1 depicts the draws from the prior as well as every 5,000th draw from the posterior distribution in two-dimensional scatter plots. We also indicate the location of the posterior mode in the 12 panels of the figure. A visual comparison of priors and posteriors suggests that the 80 observations sample contains little information on the risk-aversion parameter τ and the policy rule coefficients ψ_1 and ψ_2. There is, however, information about the degree of price-stickiness, captured by the slope coefficient κ of the Phillips-curve relationship (30). Moreover, the posteriors of steady-state parameters, the autocorrelation coefficients, and the shock standard deviations are sharply peaked relative to the prior distributions. The lack of information about some of the key structural parameters resembles the empirical findings based on actual observations.

There exists an extensive literature on convergence diagnostics for MCMC methods such as the RWM algorithm. An introduction to this literature can be found, for instance, in Robert and Casella (1999). These authors distinguish between convergence of the Markov chain to its stationary distribution, convergence of empirical averages to posterior moments, and convergence to *iid* sampling. While many convergence diagnostics are based on formal statistical tests, we will consider informal graphical methods in this paper. More specifically, we will compare draws and recursively computed means from multiple chains.

We run four independent Markov chains. Except for τ and ψ_2, all parameters are initialized at their posterior mode values. The initial values for (τ, ψ_2) are (3.0,.1), (3.0, 2.0), (.4,.1), and (.6, 2.0), respectively. As before, we set $c = 0.3$, generate 1 million draws for each chain, and plot every 5,000th draw of (τ, ψ_2) in Figure 2. Panels (1, 1) and (1, 2) of the figure depict posterior contours at the posterior mode. Visual inspection of the plots suggests that all four chains, despite the use of different starting values, converge to the same (stationary) distribution and concentrate in the high-posterior density region. Figure 3 plots recursive means for the four chains. Despite different initializations, the means converge in the long run.
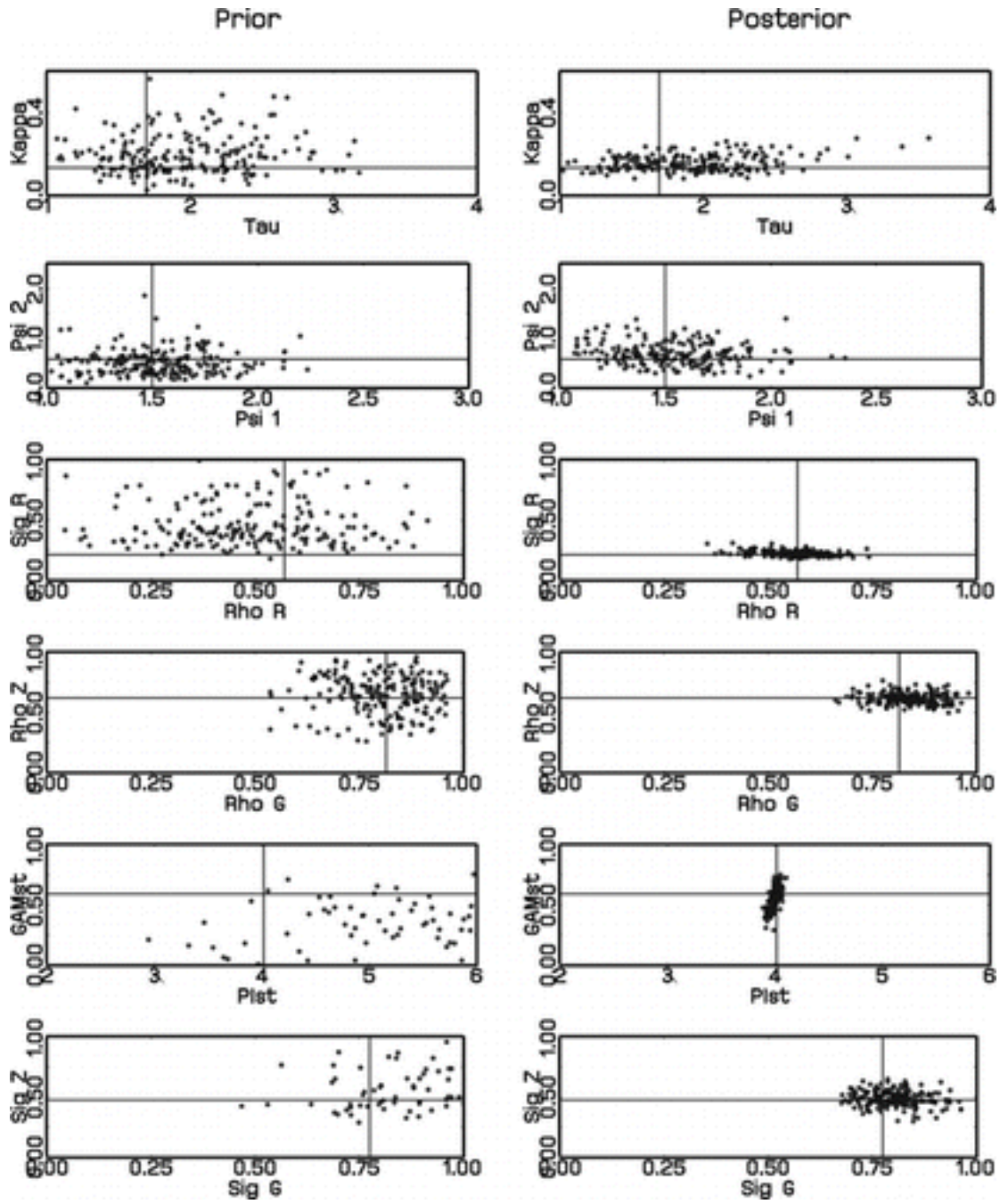
FIGURE 1 Prior and posterior – Model $\mathcal{M}_1(L)$, Data $\mathcal{D}_1(L)$. The panels depict 200 draws from prior and posterior distributions. Intersections of solid lines signify posterior mode values.
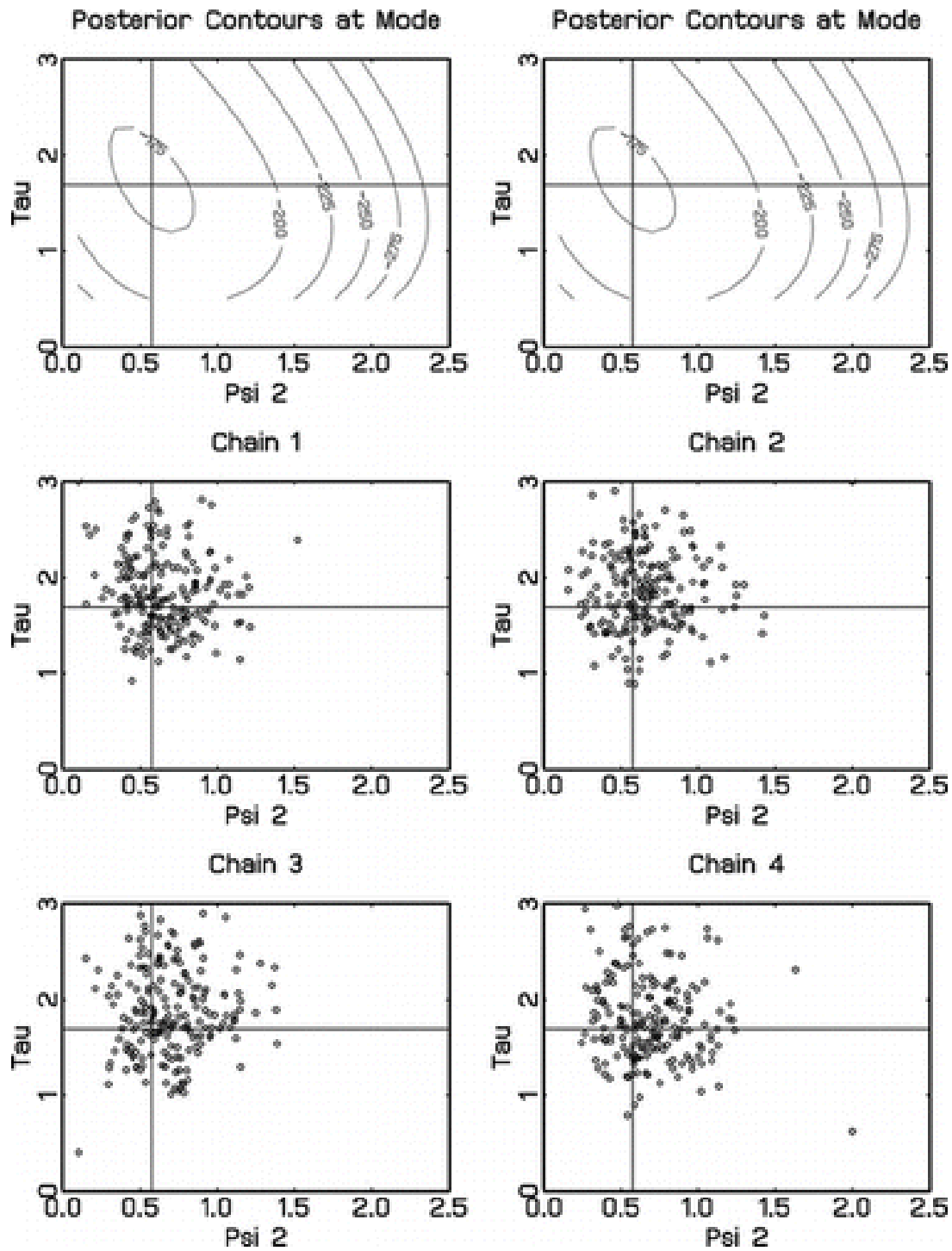
FIGURE 2 Draws from multiple chains – Model $\mathcal{M}_1(L)$, Data $\mathcal{D}_1(L)$. Panels (1, 1) and (1, 2): contours of posterior density at "low" and "high" mode as function of τ and ψ_2. Panels (2, 1) to (3, 2): 200 draws from four Markov chains generated by the Metropolis Algorithm. Intersections of solid lines signify posterior mode values.
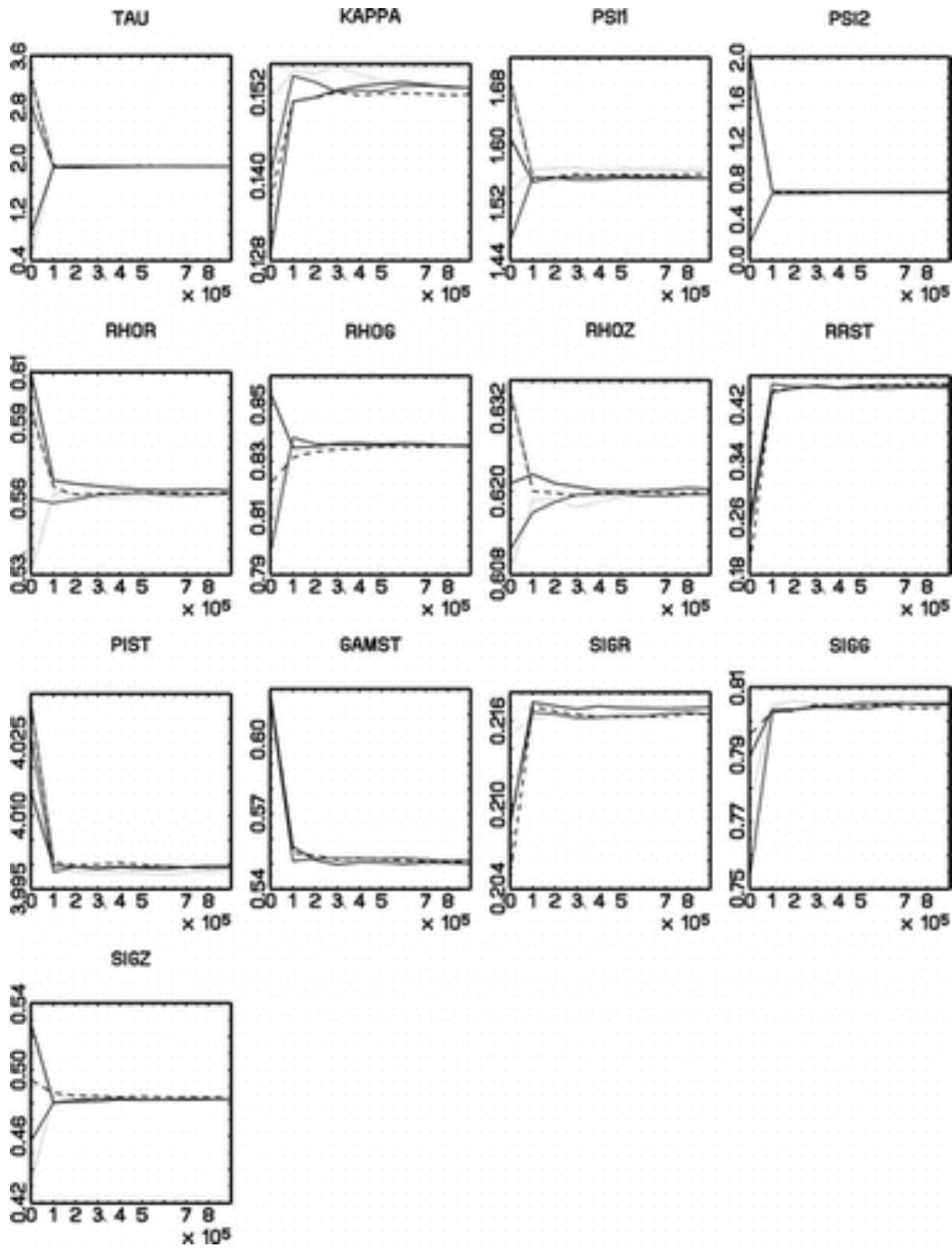
FIGURE 3 Recursive means from multiple chains – Model $\mathcal{M}_1(L)$, Data $\mathcal{D}_1(L)$. Each line corresponds to recursive means (as a function of the number of draws) calculated from one of the four Markov chains generated by the Metropolis Algorithm.

We generate another set of 1 million draws using the IS algorithm with $\nu = 3$ and $c = 1.5$. As for the draws from the RWM algorithm, we compute recursive means. Since neither the IS nor the RWM approximation of the posterior means are exact, we construct standard error estimates. For the importance sampler we follow Geweke (1999b), and for the Metropolis

chain we use Newey-West standard error estimates, truncated at 140 lags.[7] In Figure 4 we plot (recursive) confidence intervals for the RWM and the IS approximation of the posterior means. For most parameters the two confidence intervals overlap. Exceptions are, for instance, κ, $r^{(A)}$, ρ_g, and σ_g. However, the magnitude of the discrepancy is generally small relative to, say, the difference between the posterior mode and the estimated posterior means.
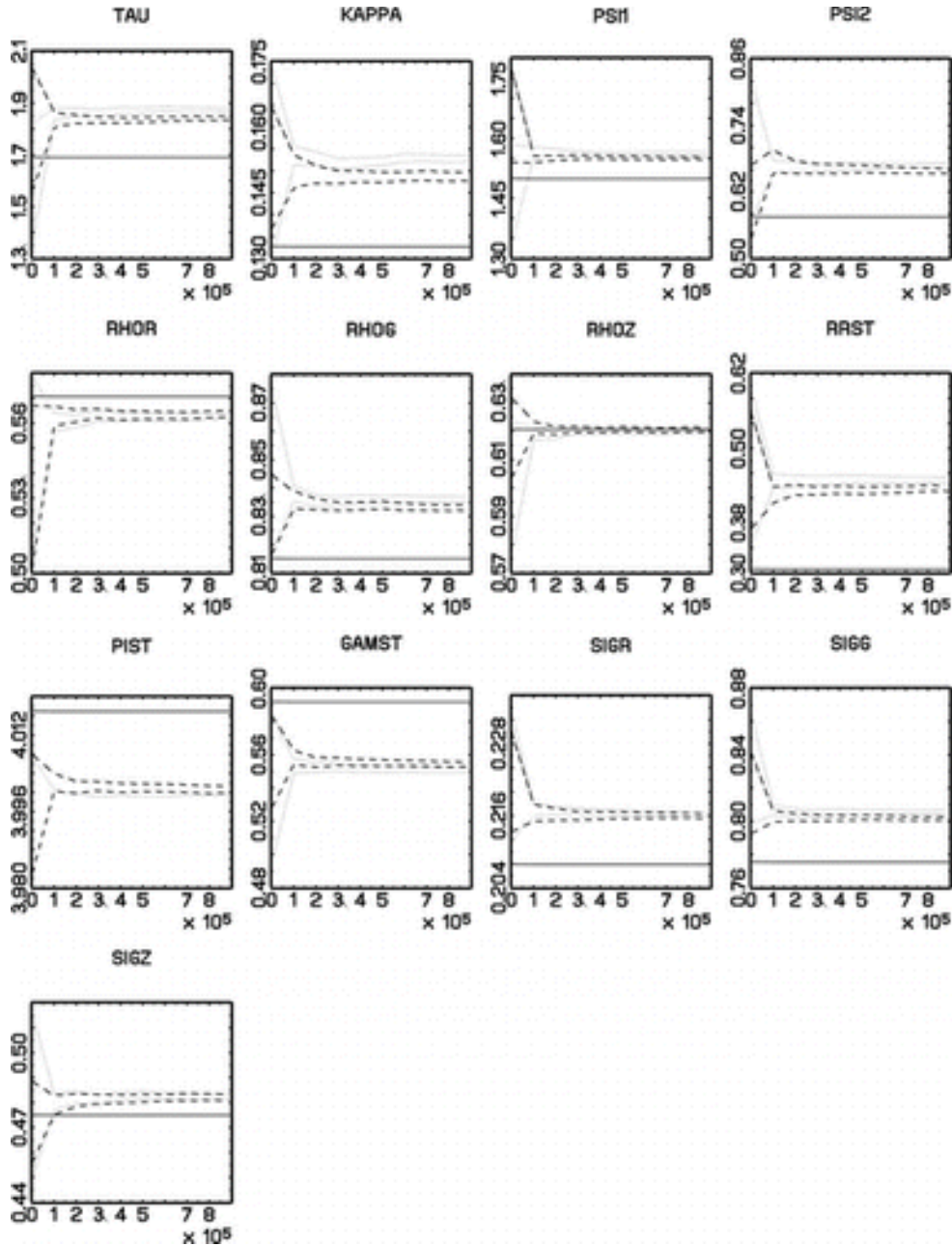

FIGURE 4 RWM algorithm vs. Importance Sampling – Model $\mathcal{M}_1(L)$, Data $\mathcal{D}_1(L)$. Panels depict posterior modes (solid), recursively computed 95% bands for posterior means based on the metropolis algorithm (dotted) and the importance sampler (dashed).

---

[7] It is not guaranteed that the Newey–West standard errors are formally valid.

4.3 A Potential Pitfall: Simulation Results for $\mathcal{M}_2(L)/\mathcal{D}_2(L)$

We proceed by estimating the output growth rule version $\mathcal{M}_2(L)$ of the DSGE model based on 80 observations generated from this model (Data Set $\mathcal{D}_2(L)$). Unlike the posterior for the output gap version, the $\mathcal{M}_2(L)$ posterior has (at least) two modes. One of the modes, which we label the "high" mode, $\bar{\theta}^{(h)}$, is located at $\tau = 2.06$ and $\psi_2 = .97$. The value of $\ln [\mathcal{L}(\theta \mid Y) p(\theta)]$ is equal to -175.48. The second ("low") mode, $\bar{\theta}^{(l)}$, is located at $\tau = 1.44$ and $\psi_2 = .81$ and attains the value -183.23. The first two panels of Figure 5 depict the contours of the non-normalized posterior density as a function of τ and ψ_2 at the low and the high mode, respectively. Panel (1, 1) is constructed by setting $\theta = \bar{\theta}^{(l)}$ and then graphing posterior contours for $\tau \in [0, 3]$ and $\psi_2 \in [0, 2.5]$, keeping all other parameters fixed at their respective $\bar{\theta}^{(l)}$ values. Similarly, Panel (1, 2) is obtained by exploring the shape of the posterior as a function of τ and ψ_2, fixing the other parameters at their respective $\bar{\theta}^{(h)}$ values. The intersection of the solid lines in panels (1, 1), (2, 1), and (3, 1) signify the low mode, whereas the solid lines in the remaining panels indicate the location of the high mode. The two modes are separated by a deep valley which is caused by complex eigenvalues of the state transition matrix.

As in Section 4.2 we generate 1 million draws each for model $\mathcal{M}_2$ from four Markov chains that were initialized at the following values for $(\tau; \psi_2)$: (3.0; .1), (3.0; 2.0), (.4; .1), and (.6; 2.0). The remaining parameters were initialized at the low (high) mode for Chains 1 and 3 (2 and 4).[8] We plot every 5,000th draw of τ and ψ_2 from the four Markov chains in Panels (2, 1) to (3, 2). Unlike in Panels (1, 1) and (1, 2), the remaining parameters are not fixed at their $\bar{\theta}^{(l)}$ and $\bar{\theta}^{(h)}$ values. It turns out that Chains 1 and 3 explore the posterior distribution locally in the neighborhood of the low mode, whereas Chains 2 and 4 move through the posterior surface near the high mode. Given the configuration of the RWM algorithm the likelihood of crossing the valley that separates the two modes is so small that it did not occur. The recursive means associated with the four chains are plotted in Figure 6. They exhibit two limit points, corresponding to the two posterior modes. While each chain appears to be stable, it only explores the posterior in the neighborhood of one of the modes.

---

[8] The covariance matrices of the proposal distributions are obtained from the Hessians associated with the respective modes.
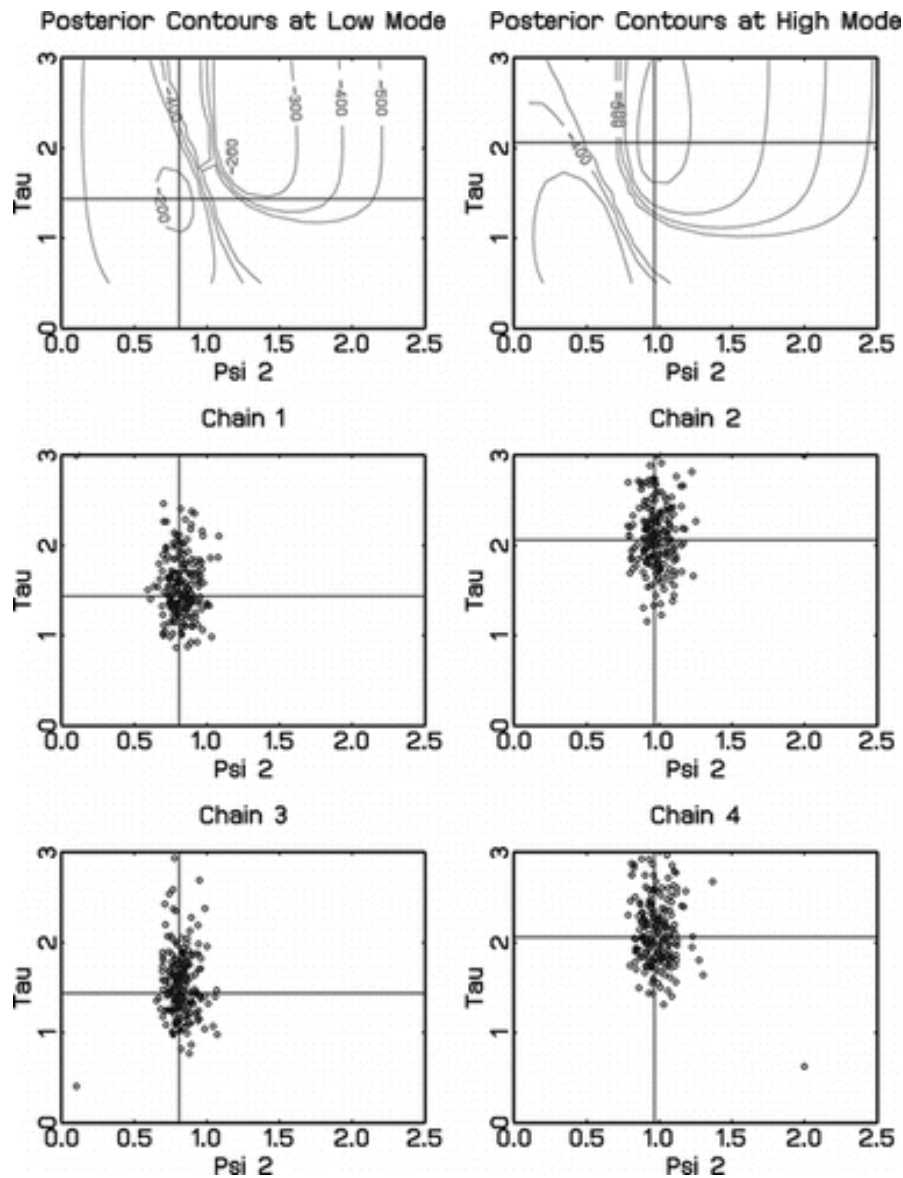
FIGURE 5 Draws from multiple chains – Model $\mathcal{M}_2(L)$, Data $\mathcal{D}_2(L)$. Panels (1, 1) and (1, 2): contours of posterior density at "low" and "high" mode as function of τ and ψ_2. Panels (2, 1) to (3, 2): 200 draws from four Markov chains generated by the metropolis algorithm. Intersections of solid lines signify "low" (left panels) and "high" (right panels) posterior mode values.
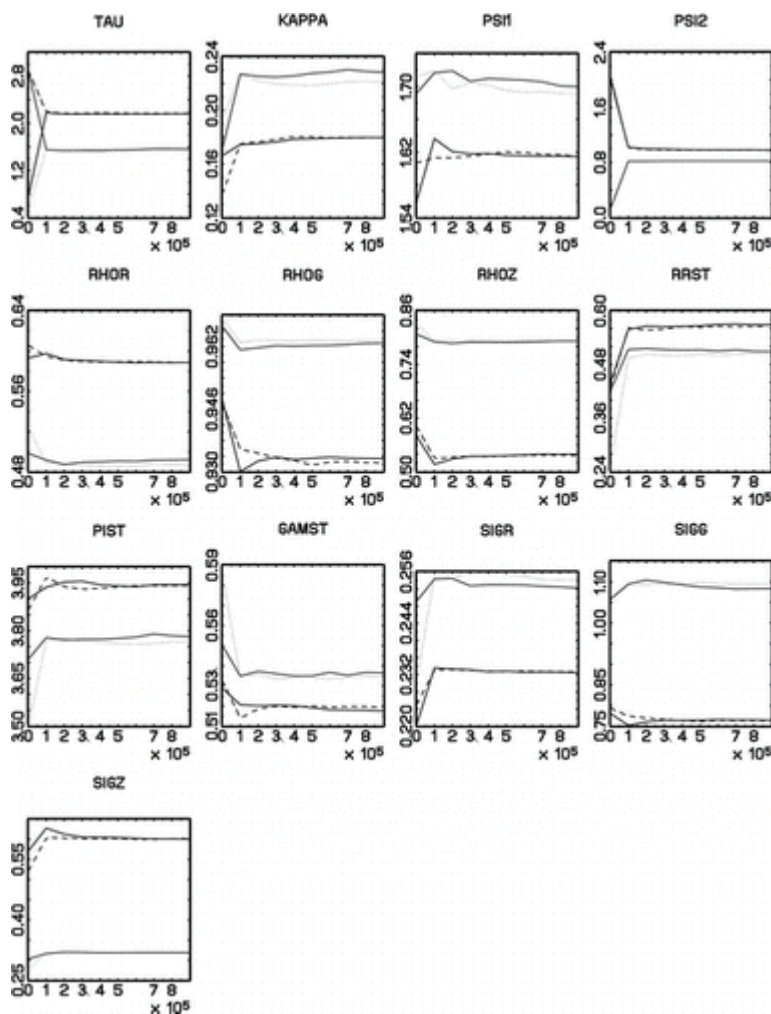
FIGURE 6 Recursive means from multiple chains – Model $\mathcal{M}_2(L)$, Data $\mathcal{D}_{(}L)$. Each line corresponds to recursive means (as a function of the number of draws) calculated from one of the four Markov chains generated by the metropolis algorithm.

We explored various modifications of the RWM algorithm by changing the scaling factor c and by setting the off-diagonal elements of $\tilde{\Sigma}$ to zero. Nevertheless, 1,000,000 draws were insufficient for the chains initialized at the low mode to cross over to the neighborhood of the high mode. Two remarks are in order. First, extremum estimators that are computed with numerical optimization methods suffer from the same problem as the Bayes estimators in this example. One might find a local rather than the global extremum of the objective function. Hence, it is good practice to start the optimization from different points in the parameter space to increase the likelihood that the global optimum is found. Similarly, in Bayesian computation it is helpful to start MCMC methods from different regions of the parameter space, or vary the distribution $q(\theta)$ in the importance sampler. Second, in many applications as well as in this example there exists a global mode that dominates the local modes. Hence an exploration of the posterior in the neighborhood of the global mode might still provide a good approximation to the overall posterior distribution. All subsequent computations are based on the output gap rule specification of the DSGE model.

4.4 Parameter Transformations for $\mathcal{M}_1(L)/\mathcal{D}_1(L)$

Macroeconomists are often interested in posterior distributions of parameter transformations $h(\theta)$ to address questions such as what fraction of the variation in output growth is caused by monetary policy shocks, and what happens to output and inflation in response to a monetary policy shock. Answers can be obtained from the moving average representation associated with the state space model composed of (33) and (38). We will focus on variance decompositions in the remainder of this subsection.

Fluctuations of output growth, inflation, and nominal interest rate in the DSGE model are due to three shocks: technology growth shocks, government spending shocks, and monetary policy shocks. Hence variance decompositions of the endogenous variables lie in a three-dimensional simplex, which can be depicted as a triangle in $\mathbb{R}^2$. In Figure 7 we plot 200 draws from the prior distribution (left panels) and 200 draws from the posterior (right panels) of the variance decompositions of output growth and inflation. The posterior draws are obtained by converting every 5,000th draw generated with the RWM algorithm. The corners Z, G, and R of the triangles correspond to decompositions in which the shocks ϵ_z, ϵ_g, and ϵ_R explain 100% of the variation, respectively.
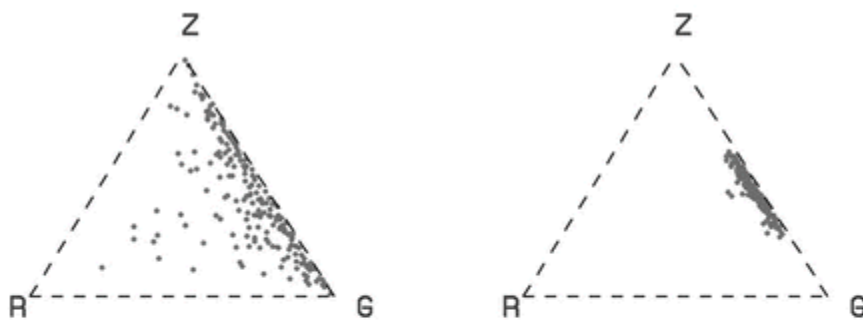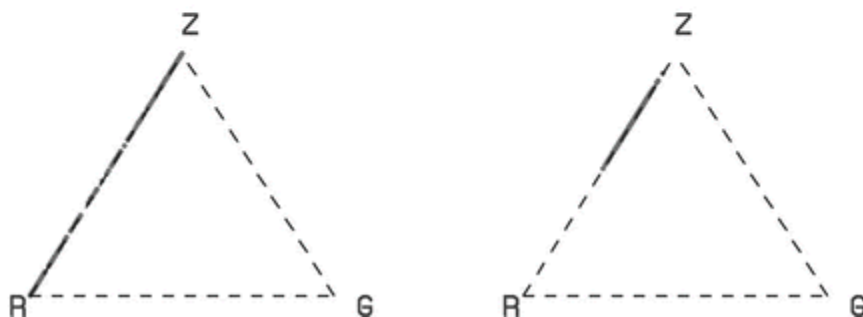

FIGURE 7 Prior and posterior variance decompositions – Model $\mathcal{M}_1(L)$, Data $\mathcal{D}_1(L)$. The panels depict 200 draws from prior and posterior distributions arranged on a 3-dimensional simplex. The three corners (Z,G,R) correspond to 100% of the variation being due to the shocks $\epsilon_{z,t}$, $\epsilon_{g,t}$, and $\epsilon_{R,t}$, respectively.

Panels (1, 1) and (2, 1) indicate that the prior distribution of the variance decomposition is informative in the sense that it is not uniform over the simplex. For instance, the output gap policy rule specification $\mathcal{M}_1$ implies that the government spending shock does not affect inflation and nominal interest rates. Hence all prior and posterior draws concentrate on the $R - Z$ edge of the simplex. While the prior mean of the fraction of inflation variability explained by the monetary policy shock is about 40%, a 90% probability interval ranges from 0 to 95%. A priori about 10% of the variation in output growth is due to monetary policy shocks. A 90% probability interval ranges from 0 to 20%. The data provide additional information about the variance decomposition. The posterior distribution is much more concentrated than the prior. The probability interval for the contribution of monetary policy shocks to output growth fluctuations shrinks to the range from 1 to 4.5%. The posterior probability interval for the fraction of inflation variability explained by the monetary policy shock ranges from 15 to 37%.

4.5. Empirical Applications

There is a rapidly growing empirical literature on the Bayesian estimation of DSGE models that applies the techniques described in this paper. The following incomplete list of references aims to give an overview of this work. Preceding the Bayesian literature are papers that use maximum likelihood techniques to estimate DSGE models. Altug (1989) estimates Kydland and Prescott's (1982) studies the macroeconomic effects of taxation in an estimated business cycle model. Leeper and Sims (1994) and Kim (2000) estimated DSGE models that are usable for monetary policy analysis.

Canova (1994), DeJong et al. (1996), and Geweke (1999a) proposed Bayesian approaches to calibration that do not exploit the likelihood function of the DSGE model and provide empirical applications that assess business cycle and asset pricing implications of simple stochastic growth models. The literature on likelihood-based Bayesian estimation of DSGE models began with work by Landon-Lane (1998), DeJong et al. (2000), Schorfheide (2000), and Otrok (2001). DeJong et al. (2000) estimate a stochastic growth model and examine its forecasting performance, Otrok (2001) fits a real business cycle with habit formation and time-to-build to the data to assess the welfare costs of business cycles, and Schorfheide (2000) considers cash-in-advance monetary DSGE models. DeJong and Ingram (2001) study the cyclical behavior of skill accumulation, whereas Chang et al. (2002) estimate a stochastic growth model augmented with a learning-by-doing mechanism to amplify the propagation of shocks. Chang and Schorfheide (2003) study the importance of labor supply shocks and estimate a home-production model. Fernández-Villaverde and Rubio-Ramírez (2004) use Bayesian estimation techniques to fit a cattle–cycle model to the data.

Variants of the small-scale New Keynesian DSGE model presented in Section 2 have been estimated by Rabanal and Rubio-Ramírez (2005a, b) for the U.S. and the Euro Area. Lubik and Schorfheide (2004) estimate the benchmark New Keynesian DSGE model without restricting the parameters to the determinacy region of the parameter space. Schorfheide (2005) allows for regime-switching of the target inflation level in the monetary policy rule. Canova (2004) estimates a small-scale New Keynesian model recursively to assess the

stability of the structural parameters over time. Galí and Rabanal (2005) use an estimated DSGE model to study the effect of technology shocks on hours worked. Large-scale models that include capital accumulation and additional real and nominal frictions along the lines of Christiano et al. (2005) have been analyzed by Smets and Wouters (2003, 2005) both for the U.S. and the Euro Area. Models similar to Smets and Wouters (2003) have been estimated by Laforte (2004), Onatski and Williams (2004), and Levin et al. (2006) to study monetary policy. Many central banks are in the process of developing DSGE models along the lines of Smets and Wouters (2003) that can be estimated with Bayesian techniques and used for policy analysis and forecasting.

Bayesian estimation techniques have also been used in the open economy literature. Lubik and Schorfheide (2003) estimate the small open economy extension of the model presented in Section 2 to examine whether the central banks of Australia, Canada, England, and New Zealand respond to exchange rates. A similar model is fitted by Del Negro (2003) to Mexican data. Justiniano and Preston (2004) extend the empirical analysis to situations of imperfect exchange rate passthrough. Adolfson et al. (2004) analyze an open economy model that includes capital accumulation as well as numerous real and nominal frictions. Lubik and Schorfheide (2006), Rabanal and Tuesta (2006), and de Walque and Wouters (2004) have estimated multicountry DSGE models.

5. MODEL EVALUATION

In Section 4 we discussed the estimation of a linearized DSGE model and reported results on the posterior distribution of the model parameters and variance decompositions. We tacitly assumed that model and prior provide an adequate probabilistic representation of the uncertainty with respect to data and parameters. This section studies various techniques that can be used to evaluate the model fit. We will distinguish the assessment of absolute fit from techniques that aim to determine the fit of a DSGE model relative to some other model. The first approach can be implemented by a posterior predictive model check and has the flavor of a classical hypothesis test. Relative model comparisons, on the other hand, are typically conducted by enlarging the model space and applying Bayesian inference and decision theory to the extended model space. Section 5.1 discusses Bayesian model checks, Section 5.2 reviews model posterior odds comparisons, and Section 5.3 describes a more elaborate model evaluation based on comparisons between DSGE models and VARs.

5.1. Posterior Predictive Checks

Predictive checks as a tool to assess the absolute fit of a probability model have been advocated, for instance, by Box (1980). A probability model is considered as discredited by the data if one observes an event that is deemed very unlikely by the model. Such model checks are controversial from a Bayesian perspective. Methods that determine whether actual data lie in the tail of a model's data distribution potentially favor alternatives that make unreasonably diffuse predictions. Nevertheless, posterior predictive checks have become a

valuable tool in applied Bayesian analysis though they have not been used much in the context of DSGE models. An introduction to Bayesian model checking can be found, for instance, in books by Gelman et al. (1995), Lancaster (2004), and Geweke (2005).

Let Y^{rep} be a sample of observations of length T that we could have observed in the past or that we might observe in the future. We can derive the predictive distribution of Y^{rep} given the current state of knowledge:

$$p(Y^{rep} \mid Y) = \int p(Y^{rep} \mid \theta) p(\theta \mid Y) d\theta. \qquad (44)$$

Let $h(Y)$ be a test quantity that reflects an aspect of the data that we want to examine. A quantitative model check can be based on Bayesian p-values. Suppose that the test quantity is univariate, non-negative, and has a unimodal density. Then one could compute probability of the tail event by

$$\int \mathcal{I}\{h(Y^{rep}) \geq h(Y)\} p(Y^{rep} \mid Y) dY^{rep}, \qquad (45)$$

where $\mathcal{I}\{x \geq a\}$ is the indicator function that is 1 if $x \geq \alpha$ and zero otherwise. A small tail probability can be viewed as evidence against model adequacy.

Rather than constructing numerical approximations of the tail probabilities for univariate functions $h(\cdot)$, we use a graphical approach to illustrate the model checking procedure in the context of model $\mathcal{M}_1(L)$. We use the following data transformations: the correlation between inflation and lagged interest rates, lagged inflation and current interest rates, output growth and lagged output growth, and output growth and interest rates. To obtain draws from the posterior predictive distribution of $h(Y^{rep})$ we take every 5,000th parameter draw of $\theta^{(s)}$ generated with the RWM algorithm, simulate a sample Y^{rep} of 80 observations[9] from the DSGE model conditional on $\theta^{(s)}$, and calculate $h(Y^{rep})$.

We consider two cases: in the case of no mis-specification, depicted in the left panels of Figure 8, the posterior is constructed based on data set $\mathcal{D}_1(L)$, whereas under mis-specification, illustrated in the two right panels of Figure 8, the posterior is obtained from data set $\mathcal{D}_5(L)$. Each panel of the figure depicts 200 draws from the posterior predictive distribution in bivariate scatter plots. Moreover, the intersections of the solid lines indicate the actual values $h(Y)$. Since $\mathcal{D}_1(L)$ was generated from model $\mathcal{M}_1(L)$, whereas $\mathcal{D}_5(L)$ was not, we would expect the actual values of $h(Y)$ to lie further in the tails of the posterior predictive distribution in the right panels than in the left panels. Indeed a visual inspection of the plots provides some evidence in which dimensions $\mathcal{M}_1(L)$ is at odds with data generated from a model that includes lags of output in the household's Euler equation. The observed autocorrelation of output growth in $\mathcal{D}_5(L)$ is much larger and the actual correlation between

---

[9] To obtain draws from the unconditional distribution of y t we initialize s 0 = 0 and generate 180 observations, discarding the first 100.

output growth and interest rates is much lower than the draws generated from the posterior predictive distribution.
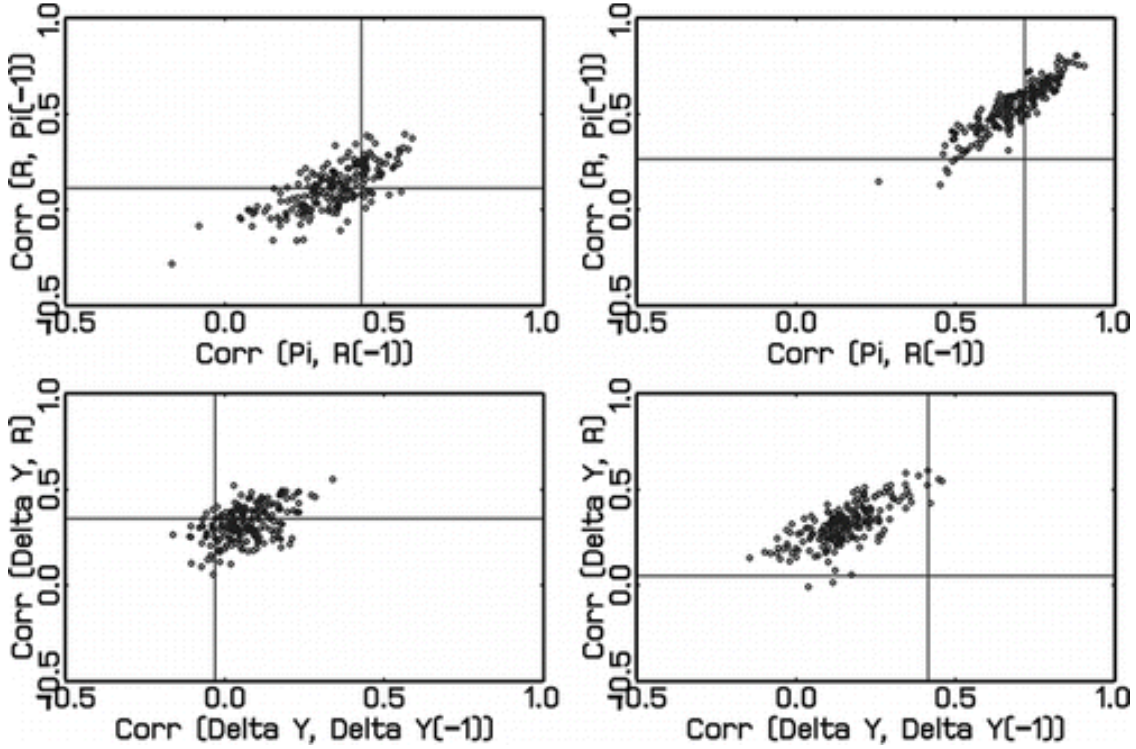

FIGURE 8 Posterior predictive check for Model $\mathcal{M}_1(L)$. We plot 200 draws from the posterior predictive distribution of various sample moments. Intersections of solid lines signify the observed sample moments.

5.2. Posterior Odds Comparisons of DSGE Models

The Bayesian framework is naturally geared toward the evaluation of relative model fit. Researchers can place probabilities on competing models and assess alternative specifications based on their posterior odds. An excellent survey on model comparisons based on posterior probabilities can be found in Kass and Rafferty (1995). For concreteness, suppose in addition to the specification $\mathcal{M}_1(L)$ we consider a version of the New Keynesian model, denoted by $\mathcal{M}_3(L)$ in which prices are nearly flexible, that is, $\kappa = 5$. Moreover, there is a model $\mathcal{M}_4(L)$, according to which the central bank does not respond to output at all and $\psi_2 = 0$. If we are willing to place prior probabilities $\pi_{i,0}$ on the three competing specifications then posterior model probabilities can be computed by

$$\pi_{i,T} = \frac{\pi_{i,0} p(Y \mid \mathcal{M}_i)}{\sum_{j=1,3,4} \pi_{j,0} p(Y \mid \mathcal{M}_j)}, \qquad i = 1, 3, 4. \qquad (46)$$

The key object in the calculation of posterior probabilities is the marginal data density $p(Y \mid \mathcal{M}_i)$, which is defined in (39).

Posterior model probabilities can be used to weigh predictions from different models. In many instances, researchers take a shortcut and use posterior probabilities to justify the

selection of a particular model specification. All subsequent analysis is then conditioned on the chosen model. It is straightforward to verify that under a 0–1 loss function (the loss attached to choosing the wrong model is one), the optimal decision is to select the highest posterior probability model. This 0–1 loss function is an attractive benchmark in situations in which the researcher believes that all the important models are included in the analysis. However, even in situations in which all the models under consideration are misspecified, the selection of the highest posterior probability model has some desirable properties. It has been shown under various regularity conditions, e.g., Phillips (1996) and Fernández-Villaverde and Rubio-Ramírez (2004), that posterior odds (or their large sample approximations) asymptotically favor the DSGE model that is closest to the 'true' data generating process in the Kullback-Leibler sense. Moreover, since the log-marginal data density can be rewritten as

$$\ln p(Y \mid \mathcal{M}) = \sum_{t=1}^{T} \ln p(y_t \mid Y^{t-1}, \mathcal{M})$$

$$= \sum_{t=1}^{T} \ln \left[\int p(y_t \mid Y^{t-1}, \theta, \mathcal{M}) p(\theta \mid Y^{t-1}, \mathcal{M}) d\theta \right], \quad (47)$$

where $\ln p(Y \mid \mathcal{M})$ can be interpreted as a predictive score (Good, 1952) and the model comparison based on posterior odds captures the relative one-step-ahead predictive performance. The practical difficulty in implementing posterior odds comparisons is the computation of the marginal data density. We will subsequently consider two numerical approaches: Geweke's (1999b) modified harmonic mean estimator and Chib and Jeliazkov's (2001) algorithm.

Harmonic mean estimators are based on the identity

$$\frac{1}{p(Y)} = \int \frac{f(\theta)}{\mathcal{L}(\theta \mid Y) p(\theta)} p(\theta \mid Y) d\theta, \quad (48)$$

where $f(\theta)$ has the property that $\int f(\theta) \, d\theta = 1$ (see Gelfand and Dey, 1994). Conditional on the choice of $f(\theta)$ an obvious estimator is

$$\hat{p}_G(Y) = \left[\frac{1}{n_{sim}} \sum_{s=1}^{n_{sim}} \frac{f(\theta^{(s)})}{\mathcal{L}(\theta^{(s)} \mid Y) p(\theta^{(s)})} \right]^{-1}, \quad (49)$$

where $\theta^{(s)}$ is drawn from the posterior $p(\theta \mid Y)$. To make the numerical approximation efficient, $f(\theta)$ should be chosen so that the summands are of equal magnitude. Geweke (1999b) proposed to use the density of a truncated multivariate normal distribution,

$$f(\theta) = \tau^{-1} (2\pi)^{-d/2} |V_\theta|^{-1/2} \exp\left[-0.5(\theta - \bar{\theta})' V_\theta^{-1} (\theta - \bar{\theta}) \right]$$

$$\times \mathcal{I} \left\{ (\theta - \bar{\theta})' V_\theta^{-1} (\theta - \bar{\theta}) \leq F_{\chi^2_d}^{-1}(\tau) \right\}. \quad (50)$$

Here $\bar{\theta}$ and V_θ are the posterior mean and covariance matrix computed from the output of the posterior simulator, d is the dimension of the parameter vector, $F_{\chi_d^2}$ is the cumulative density function of a χ^2 random variable with d degrees of freedom, and $\tau \in (0, 1)$. If the posterior of θ is in fact normal then the summands in (49) are approximately constant. Chib and Jeliazkov (2001) use the following equality as the basis for their estimator of the marginal data density:

$$p(Y) = \frac{\mathscr{L}(\theta \mid Y)p(\theta)}{p(\theta \mid Y)}. \tag{51}$$

The equation is obtained by rearranging the Bayes theorem and has to hold for all θ. While the numerator can be easily computed, the denominator requires a numerical approximation. Thus,

$$\hat{p}_{CS}(Y) = \frac{\mathscr{L}(\tilde{\theta} \mid Y)p(\tilde{\theta})}{\hat{p}(\tilde{\theta} \mid Y)}, \tag{52}$$

where we replaced the generic θ in (51) by the posterior mode $\bar{\theta}$. Within the RWM Algorithm denote the probability of moving from θ to ϑ by

$$\alpha(\theta, \vartheta \mid Y) = \min\{1, r(\theta, \vartheta \mid Y)\}, \tag{53}$$

where $r(\theta, \vartheta \mid Y)$ was in the description of the algorithm. Moreover, let $q(\theta, \bar{\theta} \mid Y)$ be the proposal density for the transition from θ to $\bar{\theta}$. Then the posterior density at the mode can be approximated by

$$\hat{p}(\tilde{\theta} \mid Y) = \frac{\frac{1}{n_{sim}}\sum_{s=1}^{n_{sim}} \alpha(\theta^{(s)}, \tilde{\theta} \mid Y)q(\theta^{(s)}, \tilde{\theta} \mid Y)}{J^{-1}\sum_{j=1}^{J} \alpha(\tilde{\theta}, \theta^{(j)} \mid Y)}, \tag{54}$$

where $\{\theta^{(s)}\}_{s=1}^{n_{sim}}$ are sampled draws from the posterior distribution with the RWM algorithm and $\{\theta^{(j)}\}_{j=1}^{J}$ are draws from $q(\bar{\theta}, \theta \mid Y)$ given the posterior mode value $\bar{\theta}$.

We first use Geweke's harmonic mean estimator to approximate the marginal data density associated with $\mathcal{M}_1(L)/\mathcal{D}_1(L)$. The results are summarized in Figure 9. We calculate marginal data densities recursively (as a function of the number of MCMC draws) for the four chains that were initialized at different starting values as discussed in Section 4. For the output gap rule all four chains lead to estimates of around -196.7. Figure 10 provides a comparison of Geweke's versus Chib and Jeliazkov's approximation of the marginal data density for the output gap rule specification. A visual inspection of the plot suggests that both estimators converge to the same limit point. However, the modified harmonic mean estimator appears to be more reliable if the number of simulations is small.
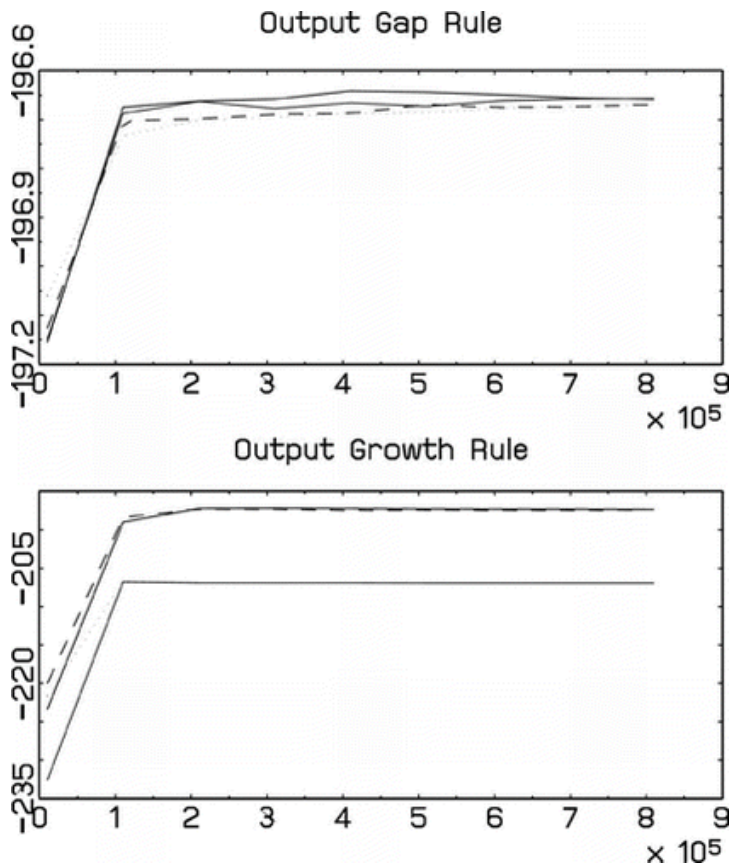
FIGURE 9 Data densities from multiple chains – Model $\mathcal{M}_1(L)$, Data $\mathcal{D}_1(L)$. For each Markov chain, log marginal data densities are computed recursively with Geweke's modified harmonic mean estimator and plotted as a function of the number of draws.
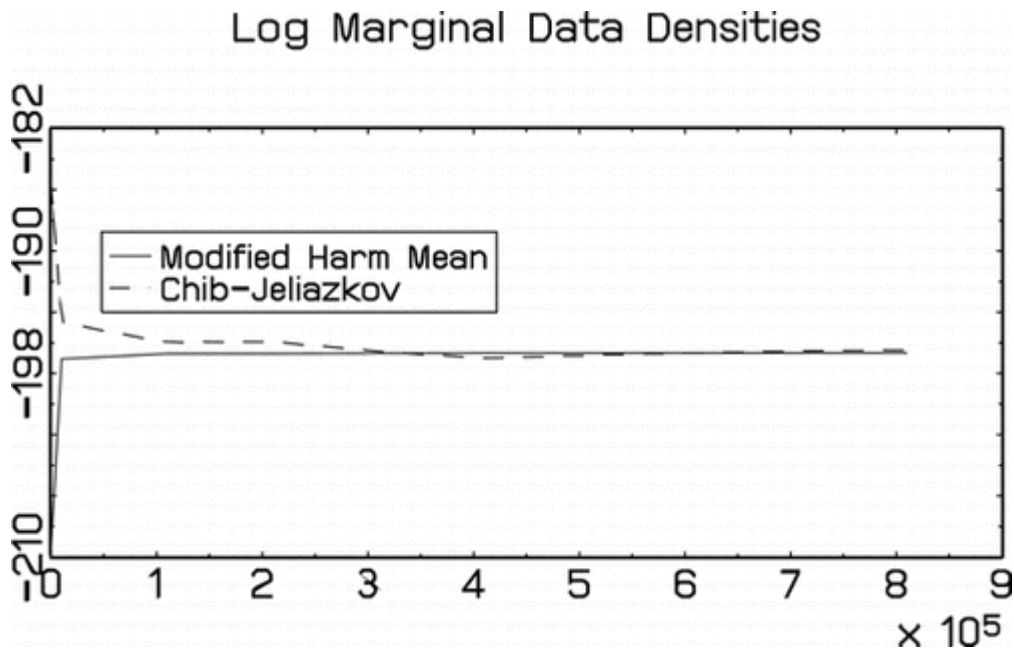

FIGURE 10 Geweke vs. Chib–Jeliazkov data densities – Model $\mathcal{M}_1(L)$, Data $\mathcal{D}_1(L)$. Log marginal data densities are computed recursively with Geweke's modified harmonic mean estimator as well as the Chib–Jeliazkov estimator and plotted as a function of the number of draws.

35

Log Marginal Data Densities

FIGURE 10 Geweke vs. Chib–Jeliazkov data densities – Model $\mathcal{M}_1(L)$, Data $\mathcal{D}_1(L)$. Log marginal data densities are computed recursively with Geweke's modified harmonic mean estimator as well as the Chib–Jeliazkov estimator and plotted as a function of the number of draws.
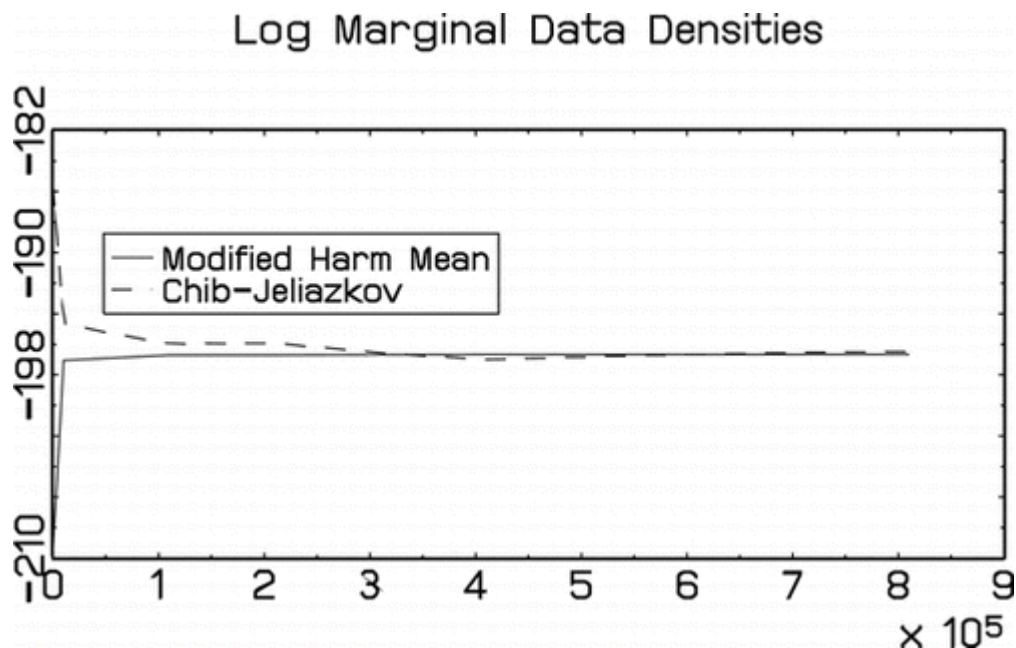
Based on data set $\mathcal{D}_1(L)$ we also estimate specification $\mathcal{M}_3(L)$ in which prices are nearly flexible and specification $\mathcal{M}_4(L)$ in which the central bank does not respond to output. The model comparison results are summarized in Table 4. The marginal data density associated with $\mathcal{M}_3(L)$ is −245.6, which translates into a Bayes factor (ratio of posterior odds to prior odds) of approximately e^{49} in favor of $\mathcal{M}_1(L)$. Hence, data set $\mathcal{D}_1(L)$ provides very strong evidence against flexible prices. Given the fairly concentrated posterior of κ depicted in Figure 1 this result is not surprising. The marginal data density of model $\mathcal{M}_4(L)$ is equal to −201.9 and the Bayes factor of model $\mathcal{M}_1(L)$ versus model $\mathcal{M}_4(L)$ is 'only' e^{5}. The DSGE model implies that when actual output is close to the target flexible price output, inflation will also be close to its target value. Vice versa, deviations of output from target coincide with deviations of inflation from its target value. This mechanism makes it difficult to identify the policy rule coefficients and imposing an incorrect value for ψ_2 is not particularly costly in terms of fit.

TABLE 4 Log marginal data densities based on $\mathcal{D}_1(L)$

Specification	$\ln p(Y \mid \mathcal{M})$	Bayes factor versus $\mathcal{M}_1(L)$
Benchmark model $\mathcal{M}_1(L)$	−196.7	1.00
Model with nearly flexible prices $\mathcal{M}_3(L)$	−245.6	exp[48.9]
No policy reaction to output $\mathcal{M}_4(L)$	−201.9	exp[5.2]

Notes: The log marginal data densities for the DSGE model specifications are computed based on Geweke's (1999a) modified harmonic mean estimator.

36

5.3. Comparison of a DSGE Model with a VAR

The notion of potential misspecification of a DSGE model can be incorporated in the Bayesian framework by including a more general reference model $\mathcal{M}_0$ into the model set. A natural choice of reference model in dynamic macroeconomics is a VAR, as linearized DSGE models, at least approximately, can be interpreted as restrictions on a VAR representation. The first step of the comparison is typically to compute posterior probabilities for the DSGE model and the VAR, which can be used to detect the presence of misspecifications. Since the VAR parameter space is generally much larger than the DSGE model parameter space, the specification of a prior distribution for the VAR parameter requires careful attention. Possible pitfalls are discussed in Sims (2003). A VAR with a prior that is very diffuse is likely to be rejected even against a misspecified DSGE model. In a more general context this phenomenon is often called Lindley's paradox. We will subsequently present a procedure that allows us to document how the marginal data density of the DSGE model changes as the cross-coefficient restrictions that the DSGE model imposes on the VAR are relaxed.

If the data favor the VAR over the DSGE model then it becomes important to investigate further the deficiencies of the structural model. This can be achieved by comparing the posterior distributions of interesting population characteristics such as impulse response functions obtained from the DSGE model and from a VAR representation that does not dogmatically impose the DSGE model restrictions. We will refer to the latter benchmark model as DSGE-VAR and discuss an identification scheme that allows us to construct structural impulse response functions for the vector autoregressive specification against which the DSGE model can be evaluated. Building on work by Ingram and Whiteman (1994) the DSGE-VAR approach of Del Negro and Schorfheide (2004) was designed to improve forecasting and monetary policy analysis with VARs. The framework has been extended to a model evaluation tool in Del Negro et al. (2006) and used to assess the fit of a variant of the Smets and Wouters (2003) model.

To construct a DSGE-VAR we proceed as follows. Consider a vector autoregressive specification of the form

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + u_t, \qquad \mathbb{E}[u_t u_t'] = \Sigma. \qquad (55)$$

Define the $k \times 1$ vector $x_t = [1, y_{t-1}', \ldots, y_{t-p}']'$, the coefficient matrix $\Phi = [\Phi_0, \Phi_1, \ldots, \Phi_p]'$, the $T \times n$ matrices Y and U composed of rows y_t' and u_t', and the $T \times k$ matrix X with rows x_t'.

Thus the VAR can be written as $Y = X\Phi + U$. Let $\mathbb{E}_\theta^D[\cdot]$ be the expectation under DSGE model and define the autocovariance matrices

$$\Gamma_{XX}(\theta) = \mathbb{E}_\theta^D[x_t x_t'], \qquad \Gamma_{XY}(\theta) = \mathbb{E}_\theta^D[x_t y_t'].$$

A VAR approximation of the DSGE model can be obtained from restriction functions that relate the DSGE model parameters to the VAR parameters,

$$\Phi^*(\theta) = \Gamma_{XX}^{-1}(\theta)\Gamma_{XY}(\theta), \qquad \Sigma^*(\theta) = \Gamma_{YY}(\theta) - \Gamma_{YX}(\theta)\Gamma_{XX}^{-1}(\theta)\Gamma_{XY}(\theta). \qquad (56)$$

This approximation is typically not exact because the state–space representation of the linearized DSGE model generates moving average terms. Its accuracy depends on the number of lags p, and the magnitude of the roots of the moving average polynomial. We will

document below that four lags are sufficient to generate a fairly precise approximation of the model $\mathcal{M}_1(L)$.[10]

In order to account for potential misspecification of the DSGE model it is assumed that there is a vector θ and matrices Φ^Δ and Σ^Δ such that the data are generated from the VAR in (55) with the coefficient matrices

$$\Phi = \Phi^*(\theta) + \Phi^\Delta, \quad \Sigma = \Sigma^*(\theta) + \Sigma^\Delta. \tag{57}$$

The matrices Φ^Δ and Σ^Δ capture deviations from the restriction functions $\Phi_*(\theta)$ and $\Sigma_*(\theta)$. Bayesian analysis of this model requires the specification of a prior distribution for the DSGE model parameters, $p(\theta)$, and the misspecification matrices. Rather than specifying a prior in terms of Φ^Δ and Σ^Δ it is convenient to specify it in terms of Φ and Σ conditional on θ. We assume

$$\Sigma \,|\, \theta \sim \mathcal{IW}\big(\lambda T \Sigma^*(\theta), \lambda T - k, n\big)$$

$$\Phi \,|\, \Sigma, \theta \sim \mathcal{N}\left(\Phi^*(\theta), \frac{1}{\lambda T}\big[\Sigma^{-1} \otimes \Gamma_{XX}(\theta)\big]^{-1}\right), \tag{58}$$

where $\mathcal{IW}$ denotes the inverted Wishart distribution.[11] The prior distribution can be interpreted as a posterior calculated from a sample of λT observations generated from the DSGE model with parameters θ; see Del Negro and Schorfheide (2004). It has the property that its density is approximately proportional to the Kullback–Leibler discrepancy between the VAR approximation of the DSGE model and the Φ-Σ VAR, which is emphasized in Del Negro et al. (2006). λ is a hyperparameter that scales the prior covariance matrix. The prior is diffuse for small values of λ and shifts its mass closer to the DSGE model restrictions as $\lambda \to \infty$. In the limit the VAR is estimated subject to the restrictions (56). The prior distribution is proper provided that $\lambda T \geq k + n$.

The joint posterior density of VAR and DSGE model parameters can be conveniently factorized as

$$p_\lambda(\Phi, \Sigma, \theta \,|\, Y) = p_\lambda(\Phi, \Sigma \,|\, Y, \theta) p_\lambda(\theta \,|\, Y). \tag{59}$$

The λ-subscript indicates the dependence of the posterior on the hyperparameter. It is straightforward to show, e.g., Zeller (1971), that the posterior distribution of Φ and Σ is also of the inverted Wishart normal form:

$$\Sigma \,|\, Y, \theta \sim \mathcal{IW}\big((1 + \lambda) T \widehat{\Sigma}_b(\theta), (1 + \lambda) T - k, n\big)$$

$$\Phi \,|\, Y, \Sigma, \theta \sim \mathcal{N}\left(\widehat{\Phi}_b(\theta), \Sigma \otimes (\lambda T \Gamma_{XX}(\theta) + X'X)^{-1}\right), \tag{60}$$

---

[10] In principle one could start from a VARMA model to avoid the approximation error. The posterior computations, however, would become significantly more cumbersome.

[11] Ingram and Whiteman (1994) constructed a VAR prior from a stochastic growth model to improve the forecast performance of the VAR. However, their setup did not allow for the computation of a posterior distribution for θ.

where $\widehat{\Phi}_b(\theta)$ and $\widehat{\Sigma}_b(\theta)$ are the given by

$$\widehat{\Phi}_b(\theta) = \left(\frac{\lambda}{1+\lambda} \Gamma_{XX}(\theta) + \frac{1}{1+\lambda} \frac{X'X}{T} \right)^{-1} \left(\frac{\lambda}{1+\lambda} \Gamma_{XY} + \frac{1}{1+\lambda} \frac{X'Y}{T} \right)$$

$$\widehat{\Sigma}_b(\theta) = \frac{1}{(1+\lambda)T} [(\lambda T \Gamma_{YY}(\theta) - Y'Y) - (\lambda T \Gamma_{YX}(\theta) + Y'X)$$

$$\times (\lambda T \Gamma_{XX}(\theta) + X'X)^{-1}(\lambda T \Gamma_{XY}(\theta) + X'Y)].$$

Hence the larger the weight λ of the prior, the closer the posterior mean of the VAR parameters is to $\Phi_*(\theta)$ and $\Sigma_*(\theta)$, the values that respect the cross-equation restrictions of the DSGE model. On the other hand, if $\lambda = (n + k)/T$, then the posterior mean is close to the OLS estimate $(X'X)^{-1} X'Y$. The marginal posterior density of θ can be obtained through the marginal likelihood

$$p_\lambda(Y \mid \theta) = \frac{|\lambda T \Gamma_{XX}(\theta) + X'X|^{-\frac{n}{2}} |(1+\lambda)T\widehat{\Sigma}_b(\theta)|^{-\frac{(1+\lambda)T-k}{2}}}{|\lambda T \Gamma_{XX}(\theta)|^{-\frac{n}{2}} |\lambda T \Sigma_*(\theta)|^{-\frac{\lambda T-k}{2}}}$$

$$\times \frac{(2\pi)^{-nT/2} 2^{\frac{n((1+\lambda)T-k)}{2}} \prod_{i-1}^{n} \Gamma[((1+\lambda)T - k + 1 - i)/2]}{2^{\frac{n(\lambda T-k)}{2}} \prod_{i-1}^{n} \Gamma[(\lambda T - k + 1 - i)/2]}.$$

A derivation is provided in Del Negro and Schorfheide (2004). The paper also shows that in large samples the resulting estimator of θ can be interpreted as a Bayesian minimum distance estimator that projects the VAR coefficient estimates onto the subspace generated by the restriction functions (56).

Since the empirical performance of the DSGE-VAR procedure crucially depends on the weight placed on the DSGE model restrictions, it is important to consider a data-driven procedure to determine λ. A natural criterion for the choice of λ in a Bayesian framework is the marginal data density

$$p_\lambda(Y) = \int p_\lambda(Y \mid \theta) p(\theta) d\theta. \tag{61}$$

For computational reasons we restrict the hyperparameter to a finite grid Λ. If one assigns equal prior probability to each grid point then the normalized $p_\lambda(Y)$'s can be interpreted as posterior probabilities for λ. Del Negro et al. (2006) emphasize that the posterior of λ provides a measure of fit for the DSGE model: high posterior probabilities for large values of λ indicate that the model is well specified and a lot of weight should be placed on its implied restrictions. Define

$$\hat{\lambda} = \underset{\lambda \in \Lambda}{\text{argmax}} \ p_\lambda(Y). \tag{62}$$

If $p_\lambda(Y)$ peaks at an intermediate value of λ, say between 0.5 and 2, then a comparison

between DSGE-VAR($\hat{\lambda}$) and DSGE model impulse responses can yield important insights about the misspecification of the DSGE model.

An impulse response function comparison requires the identification of structural shocks in the context of the VAR. So far, the VAR given in (55) has been specified in terms of reduced form disturbances u_t. According to the DSGE model, the one-step-ahead forecast errors u_t are functions of the structural shocks ϵ_t, which we represent by

$$u_t = \Sigma_{tr} \Omega \epsilon_t. \tag{63}$$

Σ_{tr} is the Cholesky decomposition of Σ, and Ω is an orthonormal matrix that is not identifiable based on the likelihood function associated with (55). Del Negro and Schorfheide (2004) proposed to construct Ω as follows. Let $A_0(\theta)$ be the contemporaneous impact of ϵ_t on y_t according to the DSGE model. Using a QR factorization, the initial response of y_t to the structural shocks can be can be uniquely decomposed into

$$\left(\frac{\partial y_t}{\partial \epsilon'_t} \right)_{DSGE} = A_0(\theta) = \Sigma^*_{tr}(\theta) \Omega^*(\theta), \tag{64}$$

where $\Sigma_{tr}*(\theta)$ is lower triangular and $\Omega*(\theta)$ is orthonormal. The initial impact of ϵ_t on y_t in the VAR, on the other hand, is given by

$$\left(\frac{\partial y_t}{\partial \epsilon'_t} \right)_{VAR} = \Sigma_{tr} \Omega. \tag{65}$$

To identify the DSGE-VAR, we maintain the triangularization of its covariance matrix Σ and replace the rotation Ω in (65) with the function $\Omega*(\theta)$ that appears in (64). The rotation matrix is chosen so that in absence of misspecification the DSGE's and the DSGE-VAR's impulse responses to all shocks approximately coincide. To the extent that misspecification is mainly in the dynamics, as opposed to the covariance matrix of innovations, the identification procedure can be interpreted as matching, at least qualitatively, the short-run responses of the VAR with those from the DSGE model. The estimation of the DSGE-VAR can be implemented as follows.

MCMC Algorithm for DSGE-VAR

1. Use the RWM algorithm to generate draws $\theta^{(s)}$ from the marginal posterior distribution $p_\lambda(\theta \mid Y)$.
2. Use Geweke's modified harmonic mean estimator to obtain a numerical approximation of $p_\lambda(Y)$.
3. For each draw $\theta^{(s)}$ generate a pair $\Phi^{(s)}$, $\Sigma^{(s)}$, by sampling from the $\mathscr{IW} - \mathscr{N}$ distribution. Moreover, compute the orthonormal matrix $\Omega^{(s)} = \Omega*(\theta)$ as described above.

We now implement the DSGE-VAR procedure for $\mathscr{M}_1(L)$ based on artificially generated data. All the results reported subsequently are based on a DSGE-VAR with $p = 4$ lags. The first step of our analysis is to construct a posterior distribution for the parameter λ. We assume that λ lies on the grid $\Lambda = \{.25, .5, .75, 1, 5, \infty\}$ and assign equal probabilities to each grid point. For $\lambda = \infty$ the misspecification matrices Φ^Δ and Σ^Δ are zero and we estimate the VAR by imposing the restrictions $\Phi = \Phi*(\theta)$ and $\Sigma = \Sigma*(\theta)$.

Table 5 reports log marginal data densities for the DSGE-VAR as a function of λ for data sets $\mathscr{D}_1(L)$ and $\mathscr{D}_5(L)$. The first row reports the marginal data density associated with the state space representation of the DSGE model, whereas the second row corresponds to the DSGE-

VAR(∞). If the VAR approximation of the DSGE model were exact, then the values in the first and second row of Table 5 would be identical. For data set $\mathcal{D}_1(L)$, which has been directly generated from the state space representation of $\mathcal{M}_1(L)$, the two values are indeed quite close. Moreover, the marginal data densities are increasing as a function of λ, indicating that there is no evidence of misspecification and that it is best to dogmatically impose the DSGE model restrictions.

TABLE 5 Log marginal data densities for $\mathcal{M}_1(L)$ DSGE-VARs

Specification	$\mathcal{D}_1(L)$	$\mathcal{D}_5(L)$
DSGE Model	-196.66	-245.62
DSGE-VAR $\lambda = \infty$	-196.88	-244.54
DSGE-VAR $\lambda = 5.00$	-198.87	-241.95
DSGE-VAR $\lambda = 1.00$	-206.57	-238.59
DSGE-VAR $\lambda = .75$	-209.53	-239.40
DSGE-VAR $\lambda = .50$	-215.06	-241.81
DSGE-VAR $\lambda = .25$	-231.20	-253.61

Notes: See Table 4.

With respect to data set $\mathcal{D}_5(L)$ the DSGE model is misspecified. This misspecification is captured in the inverse *U*-shape of the marginal data density function, which is representative for the shapes found in empirical applications in Del Negro and Schorfheide (2004) and Del Negro et al. (2006). The value $\lambda = 1.00$ has the highest likelihood. The marginal data densities drop substantially for $\lambda \geq 5$ and $\lambda \leq 0.5$. In order to assess the nature of the misspecification for $\mathcal{D}_5(L)$ we now turn to the impulse response function comparison. In principle, there are three different sets of impulse responses to be compared: responses

from the state–space representation of the DSGE model, the $\lambda = \infty$ and the $\lambda = \hat{\lambda}$ DSGE-VAR. Moreover, these impulse responses can either be compared based on the same parameter values θ or their respective posterior estimates of the DSGE model parameters. Figure 11 depicts posterior mean impulse responses for the state–space representation of the DSGE model as well as the DSGE-VAR(∞). In both cases we use the same posterior distribution of θ, namely, the one obtained from the DSGE-VAR(∞). The discrepancy between the posterior mean responses (solid and dashed lines) indicates the magnitude of the error that is made by approximating the state–space representation of the DSGE model by a fourth-order VAR. Except for the response of output to the government spending shock, the DSGE and DSGE-VAR(∞) responses are virtually indistinguishable, indicating that the approximation error is small. The (dotted) bands in Figure 11depict pointwise 90% probability intervals for the DSGE-VAR(∞) and reflect posterior parameter uncertainty with respect to θ.
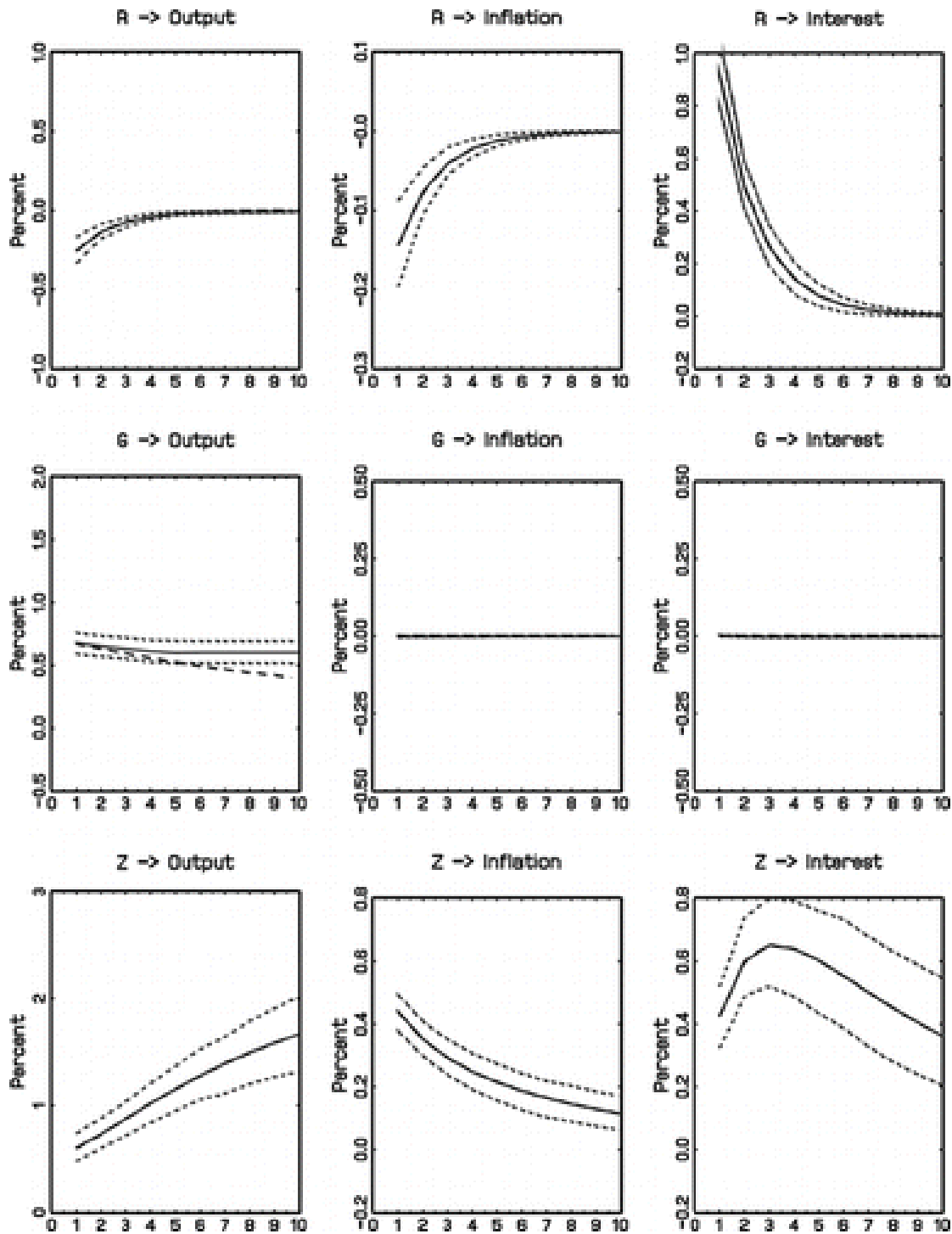
FIGURE 11 Impulse responses, DSGE, and DSGE-VAR($\lambda = \infty$) – Model $\mathcal{M}_1(L)$, Data $\mathcal{D}_5(L)$. DSGE model responses computed from state–space representation: posterior mean (solid); DSGE- VAR($\lambda = \infty$) responses: posterior mean (dashed) and pointwise 90% probability bands (dotted).

To assess the misspecification of the DSGE model we compare posterior mean responses of the $\lambda = \infty$ (dashed) and $\hat{\lambda}$ (solid) DSGE-VAR in Figure 12. Both sets of responses are constructed from the $\hat{\lambda}$ posterior of θ. The (dotted) 90% probability bands reflect uncertainty with respect to the discrepancy between the $\lambda = \infty$ and $\hat{\lambda}$ responses. A brief description of the computation can clarify the interpretation. For each draw of θ from the $\hat{\lambda}$ DSGE-VAR posterior we compute the $\hat{\lambda}$ and the $\lambda = \infty$ response functions, calculate their differences, and construct probability intervals of the differences. To obtain the dotted bands in the figure, we take the upper and lower bound of these probability intervals and shift them according to the posterior mean of the $\lambda = \infty$ response. Roughly speaking, the figure answers the question: to what extent do the $\hat{\lambda}$ estimates of the VAR coefficients deviate from the DSGE model restriction functions $\Phi_*(\theta)$ and $\Sigma_*(\theta)$. The discrepancy, however, is transformed from the Φ–Σ space into impulse response functions because they are easier to interpret.
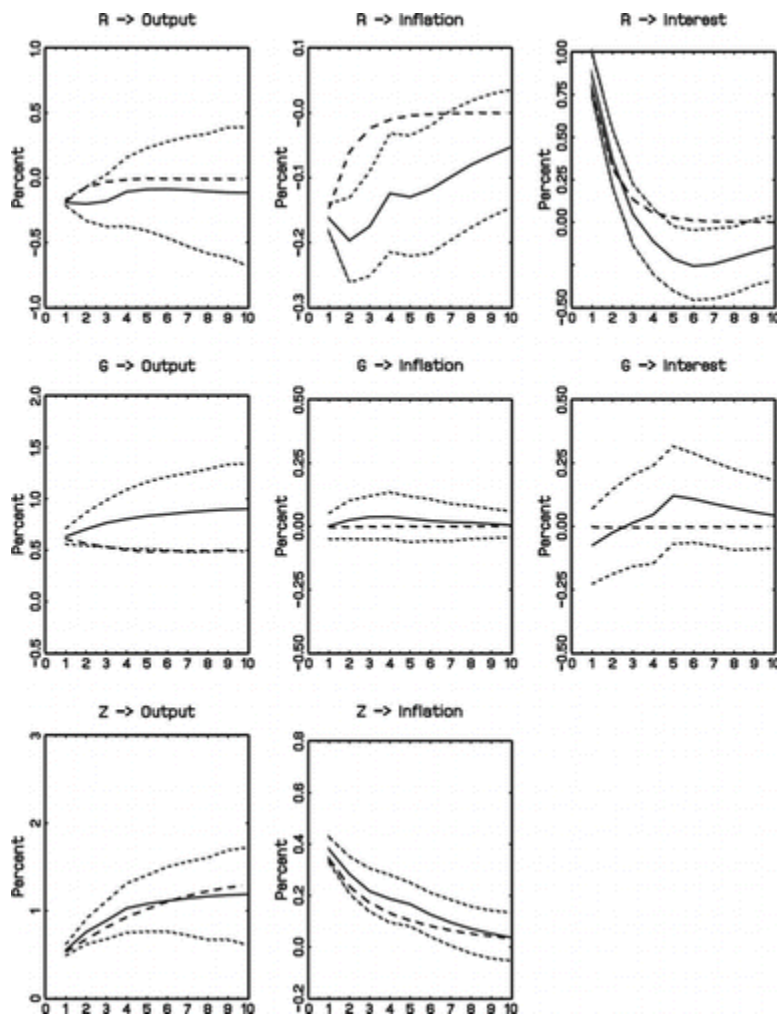

FIGURE 12 Impulse responses, DSGE-VAR($\lambda = \infty$), and DSGE-VAR($\lambda = 1$) – Model $\mathcal{M}_1(L)$, Data $\mathcal{D}_5(L)$. DSGE-VAR($\lambda = \infty$) posterior mean responses (dashed), DSGE-VAR($\lambda = 1$) posterior mean responses (solid). Pointwise 90% probability bands (dotted) signify shifted probability intervals for the difference between $\lambda = \infty$ and $\lambda = 1$ responses.

While the relaxation of the DSGE model restrictions has little effect on the propagation of the technology shock, the inflation and interest rate responses to a monetary policy shock are markedly different. According to the VAR approximation of the DSGE model, inflation returns quickly to its steady-state level after a contractionary policy shock. The DSGE-VAR ($\hat{\lambda}$) response of inflation, on the other hand, has a slight hump shape, and the reversion to the steady-state is much slower. Unlike in the DSGE-VAR(∞), the interest rate falls below steady-state in period four and stays negative for several periods.

In a general equilibrium model a misspecification of the household' decision problem can have effects on the dynamics of all the endogenous variables. Rather than looking at the dynamic responses of the endogenous variables to the structural shocks, we can also ask to what extent are the optimality conditions that the DSGE model imposes satisfied by the DSGE-VAR($\hat{\lambda}$) responses. According to the log-linearized DSGE model output, inflation, and interest rates satisfy the following relationships in response to the shock $\epsilon_{R,t}$:

$$0 = \hat{y}_t - \widehat{\mathbb{E}}_t[y_{t+1}] + \frac{1}{\tau}(\widehat{R}_t - \widehat{\mathbb{E}}_t[\pi_{t+1}]) \tag{66}$$

$$0 = \hat{\pi}_t - \beta\mathbb{E}_t[\hat{\pi}_{t+1}] - \kappa\hat{y}_t \tag{67}$$

$$\epsilon_{R,t} = \widehat{R}_t - \rho_R\widehat{R}_{t-1} - (1-\rho_R)\psi_1\hat{\pi}_t - (1-\rho_R)\psi_2\hat{y}_t \tag{68}$$

Figure 13 depicts the path of the right-hand side of Equations (66) to (68) in response to a monetary policy shock. Based on draws from the joint posterior of the DSGE and VAR parameters we can first calculate identified VAR responses to obtain the path of output, inflation, and interest rate, and in a second step use the DSGE model parameters to calculate the wedges in the Euler equation, the price setting equation, and the monetary policy rule. The dashed lines show the posterior mean responses according to the state–space representation of the DSGE model, and the solid lines depict DSGE-VAR responses. The top three panels of Figure 13 show that both for the DSGE model as well as the DSGE-VAR(∞) the three equations are satisfied. The bottom three panels of the figure are based on the DSGE-VAR($\hat{\lambda}$), which relaxes the DSGE model restrictions. While the price setting relationship and the monetary policy rule restriction seem to be satisfied, at least for the posterior mean, Panel (2, 1) indicates a substantial violation of the consumption Euler equation. This finding is encouraging for the evaluation strategy since the data set $\mathcal{D}_5(L)$ was indeed generated from a model with a modified Euler equation.
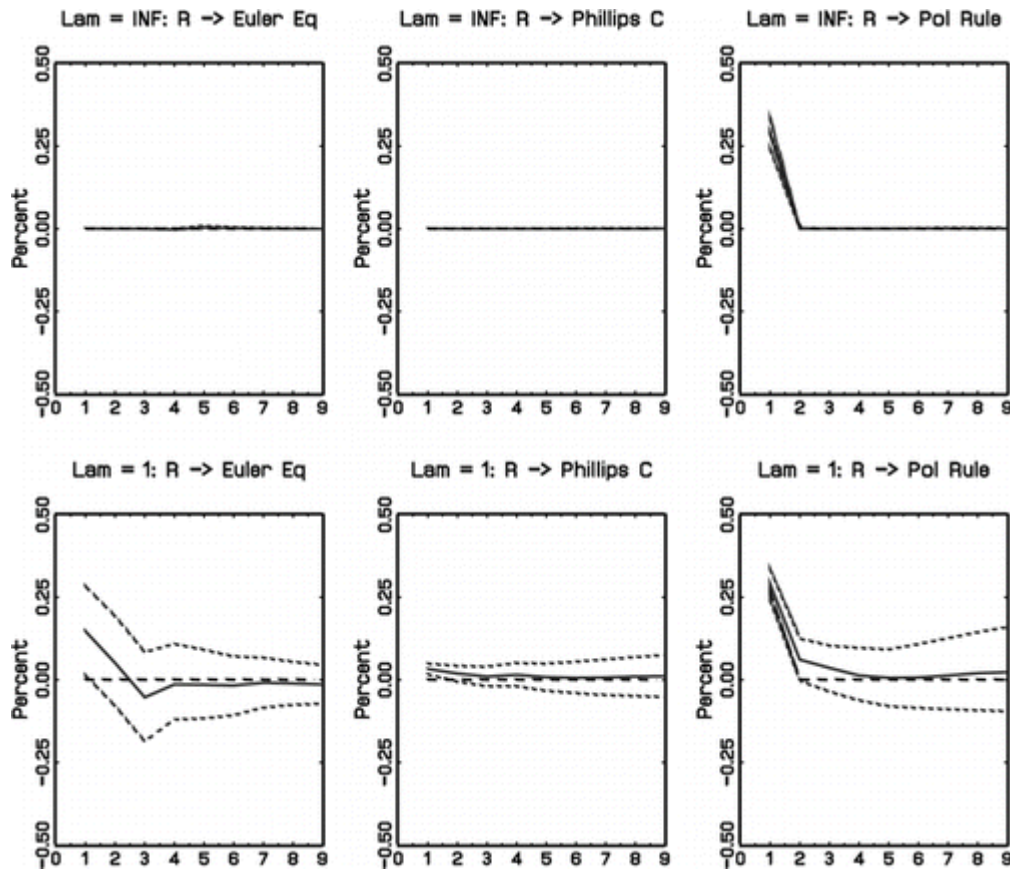
FIGURE 13 Impulse responses, DSGE-VAR($\lambda = \infty$), and DSGE-VAR($\lambda = 1$) – Model $\mathcal{M}_1(L)$, Data $\mathcal{D}_5(L)$. DSGE model responses: posterior mean (dashed); DSGE-VAR responses: posterior mean (solid) and pointwise 90% probability bands (dotted).

While this section focused mostly on the assessment of a single DSGE model, Del Negro et al. (2006) use the DSGE-VAR framework also to compare multiple DSGE models. Schorfheide (2000) proposed a loss-function-based evaluation framework for (multiple) DSGE models that augments potentially misspecified DSGE models with a more general reference model, constructs a posterior distribution for population characteristics of interest such as autocovariances and impulse responses, and then examines the ability of the DSGE models to predict the population characteristics. This prediction is evaluated under a loss function and the risk is calculated under an overall posterior distribution that averages the predictions of the DSGE models and the reference model according to their posterior probabilities. This loss-function-based approach nests DSGE model comparisons based on marginal data densities as well as assessments based on a comparison of model moments to sample moments, popularized in the calibration literature, as special cases.

6. NONLINEAR METHODS

For a non-linear/nonnormal state–space model, the linear Kalman filter cannot be used to compute the likelihood function. Instead, numerical methods have to be used to integrate the latent state vector. The evaluation of the likelihood function can be implemented with a particle filter, also known as a sequential Monte Carlo filter. Gordon et al. (1993) and Kitagawa (1996) are early contributors to this literature. Arulampalam et al. (2002) provide an excellent survey. In economics, the particle filter has been applied to analyze the stochastic volatility models by Pitt and Shephard (1999) and Kim et al. (1998). Recently Fernández-Villaverde and Rubio-Ramírez (2005) use the filter to construct the likelihood for DSGE model solved with a projection method. We follow their approach and use the particle filter for DSGE model $\mathcal{M}_1$ solved with a second-order perturbation method. A brief description of the procedure is given below. We use Y^τ to denote the $\tau \times n$ matrix with rows y_t', $t = 1,\ldots, \tau$. The vector s_t has been defined in Section 2.4.

Particle Filter

1. *Initialization*: Draw N particles s_0^i, $i = 1,\ldots,N$, from the initial distribution $p(s_0 \mid \theta)$.
   By induction, in period t we start with the particles
   $$\left\{ s_{t-1}^i \right\}_{i=1}^N, \text{ which approximate } p\left(s_{t-1} \mid Y^{t-1}, \theta\right).$$

2. Draw one-step-ahead forecasted particles $\{\tilde{s}_t^i\}_{i=1}^N$ from $p\left(s_t \mid Y^{t-1}, \theta\right)$. Note that
   $$p\left(s_t \mid Y^{t-1}, \theta\right) = \int p(s_t \mid s_{t-1}, \theta)\, p\left(s_{t-1} \mid Y^{t-1}, \theta\right) ds_{t-1} \approx \frac{1}{N} \sum_{i=1}^N p\left(s_t \mid s_{t-1}^i, \theta\right).$$
   Hence one can draw N particles from $p\left(s_t \mid Y^{t-1}, \theta\right)$ by generating one particle from $p\left(s_t \mid s_{t-1}^i, \theta\right)$ for each i.

3. *Filtering*: The goal is to approximate
   $$p\left(s_t \mid Y^t, \theta\right) = \frac{p\left(y_t \mid s_t, \theta\right)\, p\left(s_t \mid Y^{t-1}, \theta\right)}{p\left(y_t \mid Y^{t-1}, \theta\right)}, \qquad (69)$$
   which amounts to updating the probability weights assigned to the particles $\{\tilde{s}_t^i\}_{i=1}^N$. We begin by computing the non-normalized importance weights $\tilde{\pi}_t^i = p\left(y_t \mid \tilde{s}_t^i, \theta\right)$.
   The denominator in (69) can be approximated by
   $$p\left(y_t \mid Y^{t-1}, \theta\right) = \int p\left(y_t \mid s_t, \theta\right) p\left(s_t \mid Y^{t-1}, \theta\right) ds_t \approx \frac{1}{N} \sum_{i=1}^N \tilde{\pi}_t^i. \qquad (70)$$

46

Now define the normalized weights

$$\pi_t^i = \frac{\tilde{\pi}_t^i}{\sum_{j=1}^{N} \tilde{\pi}_t^j}$$

and note that the importance sampler $\left\{\tilde{s}_t^i, \pi_t^i\right\}_{i=1}^{N}$ approximates[12] the updated density $p\left(s_t \mid Y^t, \theta\right)$.

4.
   Resampling: We now generate a new set of particles $\left\{s_t^i\right\}_{i=1}^{N}$ by resampling with replacement N times from an approximate discrete representation of $p\left(s_t \mid Y^t, \theta\right)$ given by $\left\{\tilde{s}_t^i, \pi_t^i\right\}_{i=1}^{N}$ so that

$$\Pr\left(s_t^i = \tilde{s}_t^i\right) = \pi_t^i, \qquad i = 1, \dots, N.$$

The resulting sample is in fact an *iid* sample from the discretized density of $p(s_t \mid Y^t, \theta)$ and hence is equally weighted.

5. *Likelihood Evaluation*: According to (70) the log likelihood function can be approximated by

$$\ln \mathscr{L}(\theta \mid Y^T) = \ln p(y_1 \mid \theta) + \sum_{t=2}^{T} \ln p\left(y_t \mid Y^{t-1}, \theta\right) \approx \sum_{t=1}^{N} \left(\frac{1}{N} \sum_{i=1}^{N} \tilde{\pi}_t^i\right).$$

To compare results from first and second-order accurate DSGE model solutions we simulated artificial data of 80 observations from the quadratic approximation of model $\mathcal{M}_1$ with parameter values reported in Table 3 under the heading $\mathcal{D}_1(Q)$. In the remainder of this section we will refer to these parameters as "true" values as opposed to "pseudotrue" values to be defined later. These true parameter values by and large resemble the parameter values that have been used to generate data from the log-linear DSGE model specifications. However, there are some exceptions. The degree of imperfect competition, ν, is set to 0.1, which implies a steady-state markup of 11% that is consistent with the estimates of Basu (1995). $1/g$ is chosen as .85, which implies that the steady-state consumption amounts to 85% of output. The slope coefficient of the Phillips curve, κ, is chosen to be .33, which implies less price stickiness than in the linear case. The implied quadratic adjustment cost coefficient, ϕ, is 53.68. We also make monetary policy less responsive to the output gap by setting $\psi_2 = .125$. $\gamma^{(Q)}$, $r^{(A)}$, and $\pi^{(A)}$ are chosen as .55, 1.0, and 3.2 to match the average output growth rate, interest rate, and inflation between the artificial data and the real U.S. data. These values are different from the mean of the historical observations because in the non-linear version of the DSGE model means differ from steady states. Other exogenous process

---

[12] The sequential importance sampler (SIS) is a Monte Carlo method which utilized this aspect, but it is well documented that it suffers from a degeneracy problem, where after a few iterations, all but one particle will have negligible weight.

parameters, $\rho_g, \rho_z, \sigma_R, \sigma_g,$ and $\sigma_z,$ are chosen so that the second moments are matched to the U.S. data. For computational convenience, a measurement error is introduced to each measurement equation, whose standard deviation is set as 20% of that of each observation.[13]

6.1 Configuration of the Particle Filter

There are several issues concerning the practical implementation of the particle filter. First, we need a scheme to draw from the initial state distribution, $p(s_0 \mid \theta)$. In the linear case, it is straightforward to calculate the unconditional distribution of s_t associated with the vector autoregressive representation (33). For the second-order approximation we rewrite the state transition equation (37) as follows. Decompose $s_t = [x_t', \xi_t']'$, where ξ_t is composed of the exogenous processes $\hat{\epsilon}_{R,t}, \hat{g}_t,$ and $\hat{z}_t.$ The law of motion for the endogenous state variables $x_{j,t}$ can be expressed as

$$x_{j,t} = \Psi_j^{(0)} + \Psi_j^{(1)} w_t + w_t' \Psi_j^{(2)} w_t \tag{71}$$

where $w_t = [x_{t-1}', \xi_t']'$. We generate x_0 by drawing ξ_0 from its unconditional distribution (the law of motion for ξ_t is linear) and setting $x_{-1} = x$, where x is the steady-state of x_t.

Second, we have to choose the number of particles. For a good approximation of the prediction error distribution, it is desirable to have many particles, especially enough particles to capture the tails of $p(s_t \mid Y^t)$. Moreover, the number of particles affects the performance of the resampling algorithm. If the number of particles is too small, the resampling will not work well. In our implementation, the stratified resampling scheme proposed by Kitagawa (1996) is applied. It is optimal in terms of variance in the class of unbiased resampling schemes. If the measurement errors in the conditional distribution $p(y_t \mid s_t)$ are small, more particles are needed to obtain an accurate approximation of the likelihood function and to ensure that the posterior weights π_t^i do not assign probability one to a single particle. In our application, we found that 40,000 particles were enough to get stable evaluations of the likelihood given the size of the measurement errors.

Even though the estimation procedure involves extensive random sampling, the particle filter can be readily implemented on a good desktop computer. We implement most of the procedure in MATLAB (2005) so that we can exploit the symbolic toolbox to solve the DSGE model, but it takes too much time to evaluate the likelihood function using the particle filter. The filtering step is implemented as FORTRAN mex library, so we can call it as a function in MATLAB. Based on a sample of 80 observations we can generate 1,000 draws with 40,000 particles in 1 h and 40 min (6.0 sec per draw) on the AMD64 3000+ machine with 1 GB RAM. Linear methods are more than 200 times faster: 1,000 draws are generated in 24 sec in our Matlab routines and 3 sec in our GAUSS routines.

---

[13] Without the measurement error two complications arise: since $p(y_t \mid s_t)$ degenerates to a point-mass and the distribution of s_t is that of a multivariate noncentral $\chi2$ the evaluation of $p(y_t \mid Y_{t-1})$ becomes difficult. Moreover, the computation of $p(s_t \mid Y_t)$ requires the solution of a system of quadratic equations.

6.2. A Look at the Likelihood Function

We will begin our analysis by studying profiles of likelihood functions. For brevity, we will refer to the likelihood obtained from the first-order (second-order) accurate solution of the DSGE model as linear (quadratic) likelihood. It is natural to evaluate the quadratic likelihood function in the neighborhood of the true parameter values given in Table 3. However, we evaluate the linear likelihood around the pseudo-true parameter values, which are obtained by finding the mode of the linear likelihood using 3,000 observations. The parameter values for ν and $1/g$ are fixed at their true values because they do not enter the linear likelihood. Recall that we had introduced a reduced-form Phillips curve parameter κ in (32) because ν and ϕ cannot be identified with the log-linearized DSGE model.

Figure 14 shows the log-likelihood profiles along each structural parameter dimension. Two features are noticeable in this comparison. First, the quadratic log-likelihood peaks around the true parameter values, while the linear log-likelihood attains its maximum around the pseudo-true parameter values. The differences between the peaks are most pronounced for the steady-state parameters $r^{(A)}$ and $\pi^{(A)}$. While in the linearized model steady states and long-run averages coincide, they differ if the model is solved by a second-order approximation. For some parameters the quadratic log-likelihood function is more concave than the linear one, which implies that the non-linear approach is able to extract more information on the structural parameters from the data. For instance, it appears that the monetary policy parameter such as ψ_1 can be more precisely estimated with the quadratic approximation.

As noted before, the quadratic approximation can identify some structural parameters that are unidentifiable under the log-linearized model. g does not enter the linear version of the model and hence the log-likelihood profile along this dimension is flat. In the quadratic approximation, however, the log-likelihood is slightly sloped in $1/g = c/y$ dimension. Moreover, the linear likelihood is flat for the values of ν and ϕ that imply a constant κ. This is illustrated in Figure 15, which depicts the linear and quadratic likelihood contours in ν-ϕ space. The contours of the linear likelihood trace out iso-κ lines. The right panel of Figure 15 indicates that the iso-κ lines intersect with the contours of the quadratic likelihood function, which suggests that ν and ϕ are potentially separately identifiable. However, our sample of 80 observations is too short to obtain a sharp estimate of the two parameters.
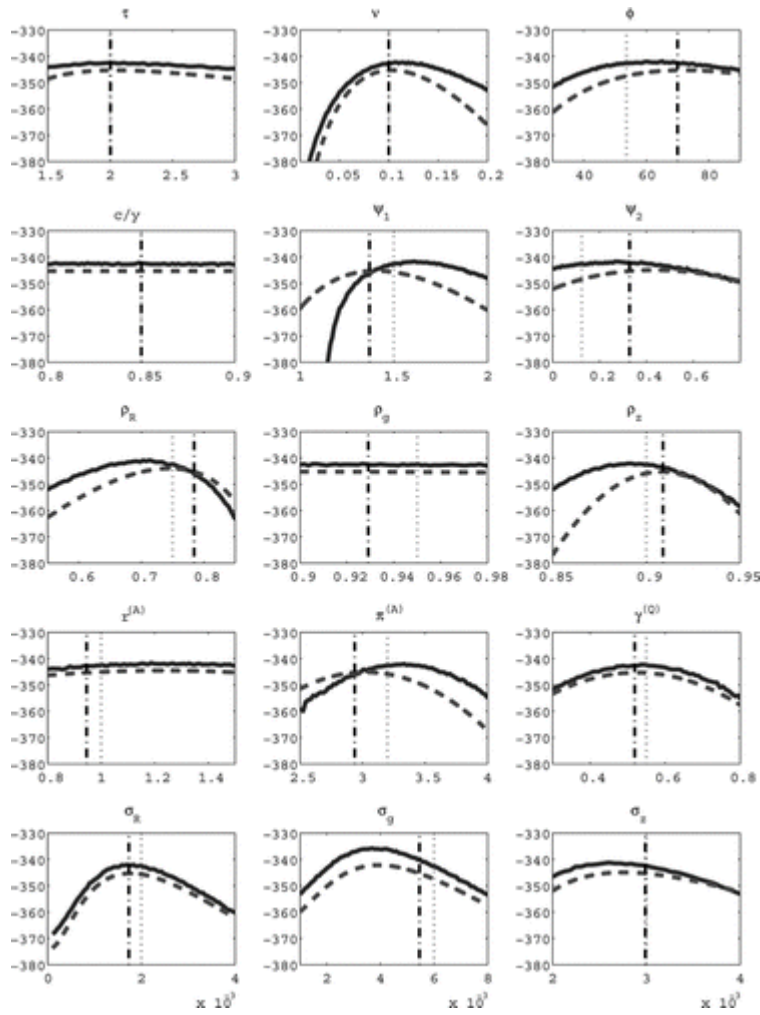
FIGURE 14 Linear vs. quadratic approximations: likelihood profiles. Data Set $\mathcal{D}_1(Q)$. Likelihood profile for each parameter: $\mathcal{M}_1(L)$/Kalman filter (dashed) and $\mathcal{M}_1(Q)$/particle filter (solid). 40,000 particles are used. Vertical lines signify true (dotted) and pseudo-true (dashed-dotted) values.
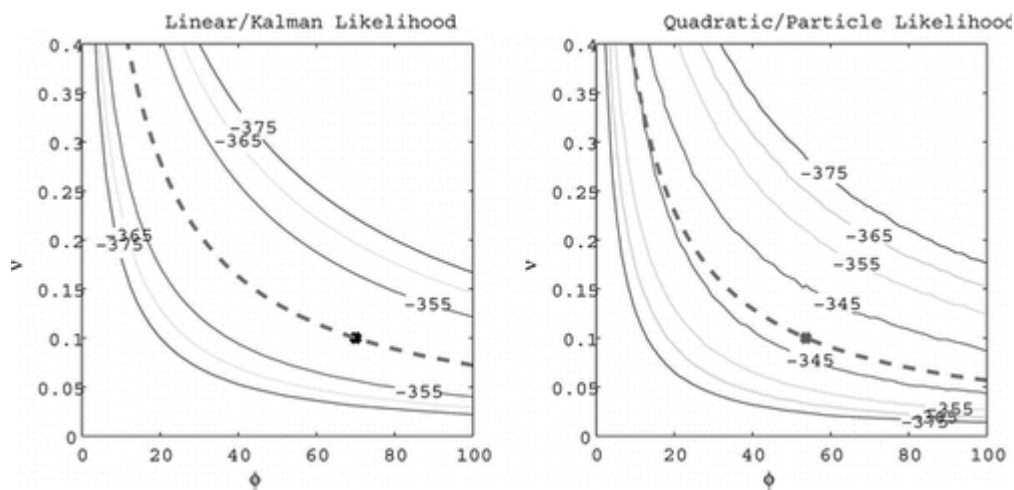

FIGURE 15 Linear vs. quadratic approximations: Likelihood contours. Data set $\mathcal{D}_1(Q)$. Contours of likelihood (solid) for $\mathcal{M}_1(L)$ and $\mathcal{M}_1(Q)$. 40,000 particles are used. Large dot indicates true and pseudotrue parameter value. κ is constant along dashed lines.

6.3. The Posterior

We now compare the posterior distributions obtained from the linear and non-linear analysis. In both cases we use the same prior distribution reported in Table 3. Unlike in the analysis of the likelihood function we now substitute the adjustment cost parameter ϕ with the Phillips-curve parameter κ in the quadratic approximation of the DSGE model. A beta prior with mean .1 and standard deviation 0.05 is placed on ν, which roughly covers micro evidences of 10–15% markup. Note that $1 - 1/g$ is the steady-state government spending–output ratio, and hence a beta prior with mean 0.85 and standard deviation .1 is used for $1/g$.

We use the RWM algorithm to generate draws from the posterior distributions associated with the linear and quadratic approximations of the DSGE model. We first compute the posterior mode for the linear specification with the additional parameters, ν and $1/g$, fixed at their true values. After that, we evaluate the Hessian to be used in the RWM algorithm at the (linear) mode without fixing ν and $1/g$, so that the inverse Hessian reflects the prior variance of the additional parameters. This Hessian is used for both the linear and the non-linear analysis. We use scaling factors .5 (linear) and .25 (quadratic) to target an acceptance rate of about 30%. The Markov chains are initialized in the neighborhood of the pseudo-true and true parameter values, respectively.

Figures 16 and 17 depict the draws from prior, linear posterior, and quadratic posterior distribution. 100,000 draws from each distribution are generated and every 500th draw is plotted. The draws tend to be more concentrated as we move from prior to posterior. While, for most of our parameters, linear and quadratic posteriors are quite similar, there are a few exceptions. The quadratic posterior mean of ψ_1 is larger than the linear posterior mean and closer to the true value. The quadratic posterior means for the steady-state parameters $r^{(A)}$ and $\pi^{(A)}$ are also greater than the corresponding linear posterior means, which is consistent with the positive difference between "true" and "pseudotrue" values of these parameters.

Now consider the parameter ν that determines the demand elasticity for the differentiated products. Conditional on κ this parameter is not identifiable under the linear approximation and hence its posterior is not updated. The parameter does, however, affect the second-order terms that arise in the quadratic approximation of the DSGE model. Indeed, the quadratic posterior of ν is markedly different from the prior as it concentrates around 0.05. Moreover, there is a clear negative correlation between κ and ν according to the quadratic posterior. Table 6 reports the log marginal data densities for our linear ($\mathcal{M}_1(L)$) and quadratic ($\mathcal{M}_1(Q)$) approximations of the DSGE model based on Data Set $\mathcal{D}_1$ (Q). The log marginal likelihoods are −416.06 for $\mathcal{M}_1(L)$ and −408.09 for $\mathcal{M}_1(Q)$, which imply a Bayes factor of about e^8 in favor of the quadratic approximation. This result suggests that $\mathcal{D}_1(Q)$ exhibits nonlinearities that can be detected by the likelihood-based estimation methods. Our simulation results with respect to linear versus non-linear estimation are in line with the findings reported in Fernández-Villaverde and Rubio-Ramírez (2005) for a version of the neoclassical stochastic growth model. In their analysis the non-linear solution combined with the particle filter delivers a substantially better fit of the model to the data as measured by the marginal data density, both for simulated as well as actual data. Moreover, the authors point out that the differences in terms of point estimates, although relatively small in magnitude, may have important effects on the moments of the model.
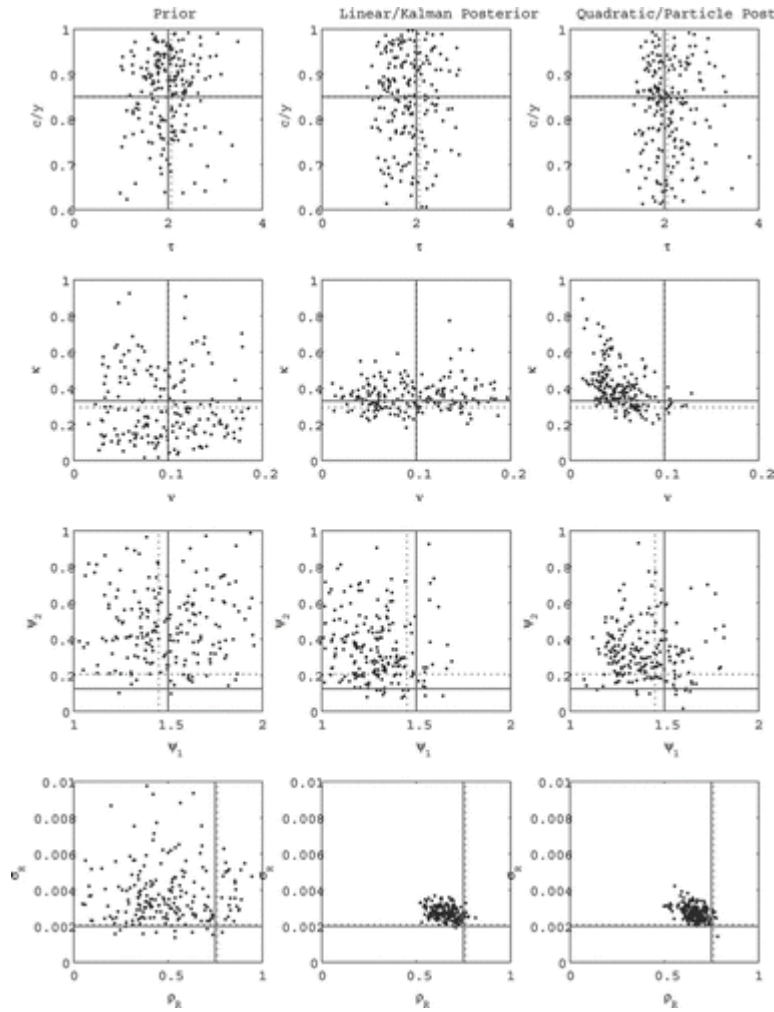
FIGURE 16 Posterior draws: linear vs. quadratic approximation I. Data set $\mathcal{D}_1(Q)$. 100,000 draws from the prior and posterior distributions. Every 500th draw is plotted. Intersections of solid and dotted lines signify true and pseudotrue parameter values. 40,000 particles are used.
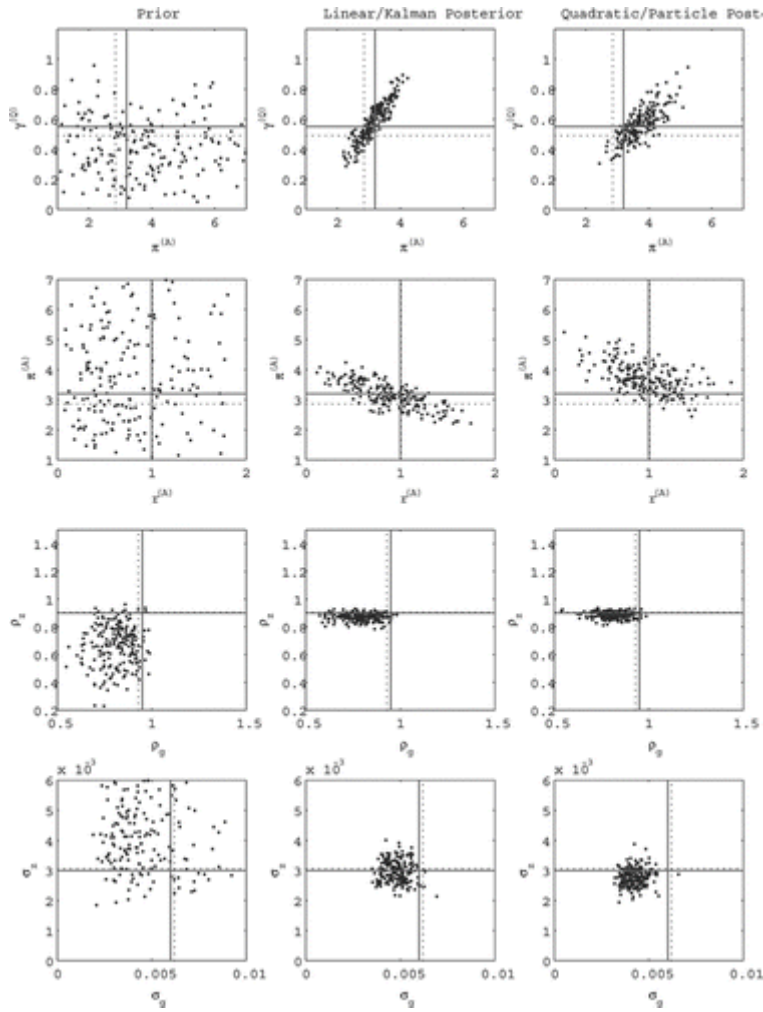
FIGURE 17 Posterior draws: Linear vs. quadratic approximation II. See Figure 16.

TABLE 6 Log marginal data densities based on $\mathcal{D}_1(Q)$

Specification	$\ln p(Y \mid \mathcal{M}_1)$	Bayes factor vesus $\mathcal{M}_1(Q)$
Benchmark model, linear solution $\mathcal{M}_1(L)$	−416.06	exp[7.97]
Benchmark model, quadratic solution $\mathcal{M}_1(Q)$	−408.09	1.00

Notes: See Table 4.

7. CONCLUSIONS AND OUTLOOK

There exists by now a large and growing body of empirical work on the Bayesian estimation and evaluation of DSGE models. This paper has surveyed the tools used in this literature and illustrated their performance in the context of artificial data. While most of the existing empirical work is based on linearized models, techniques to estimate DSGE models solved with non-linear methods have become implementable thanks to advances in computing technology. Nevertheless, many challenges remain. Model size and dimensionality of the parameter space pose a challenge for MCMC methods. Lack of identification of structural parameters is often difficult to detect since the mapping from the structural form of the DSGE model into a state–space representation is highly non-linear and creates a challenge for scientific reporting as audiences are typically interested in disentangling information provided by the data from information embodied in the prior. Model misspecification is and will remain a concern in empirical work with DSGE models despite continuous efforts by macroeconomists to develop more adequate models. Hence it is important to develop methods that incorporate potential model misspecification in the measures of uncertainty constructed for forecasts and policy recommendations.

REFERENCES

1. Adolfson, M., Laséen, S., Lindé, J., Villani, M. (2004). Bayesian estimation of an open economy DSGE model with incomplete pass-through. Journal of International Economics, forthcoming .

2. Altug, S. (1989). Time-to-build and aggregate fluctuations: Some new evidence. International Economic Review 30 (4): 889 – 920.

3. Anderson, G. (2000). A reliable and computationally efficient algorithm for imposing the saddle point property in dynamic models. Manuscript, Federal Reserve Board of Governors.

4. Arulampalam, M. S., Maskell, S., Gordon, N., Clapp, T. (2002). A tutorial on particle filters for online non-linear/non-Gaussian Bayesian tracking. IEEE Transactions on Signal Processing 50 (2): 173 – 188.

5. Aruoba, S. B., Fernández-Villaverde, J., Rubio-Ramírez, J. F. (2004). Comparing solution methods for dynamic equilibrium economies. Journal of Economic Dynamics and Control 30 (12): 2477 – 2508.

6. Basu, S. (1995). Intermediate goods and business cycles: implications for productivity and welfare. American Economic Review 85 (3): 512 – 531.

7. Beyer, A., Farmer, R. (2004). On the indeterminacy of new Keynesian economics. ECB Working Paper 323.

8. Bierens, H. J. (2007). Econometric analysis of linearized singular dynamic stochastic general equilibrium models. Journal of Econometrics, forthcoming 136 (2): 595 – 627.

9. Bils, M., Klenow, P. (2004). Some evidence on the importance of sticky prices. Journal of Political Economy 112 (5): 947 – 985.

10. Binder, M., Pesaran, H. (1997). Multivariate linear rational expectations models: characterization of the nature of the solutions and their fully recursive computation. Econometric Theory 13 (6): 877 – 888.

11. Blanchard, O. J., Kahn, C. M. (1980). The solution of linear difference models under rational expectations. Econometrica 48 (5): 1305 – 1312.

12. Box, G. E. P. (1980). Sampling and Bayes inference in scientific modelling and robustness. Journal of the Royal Statistical Society Series A 143 (4): 383 – 430.

13. Canova, F. (1994). Statistical inference in calibrated models. Journal of Applied Econometrics 9: S123 – S144.

14. Canova, F. (2004). Monetary policy and the evolution of the US economy: 1948-2002. Manuscript, IGIER and Universitat Pompeu Fabra.

15. Canova, F. (2007). Methods for Applied Macroeconomic Research. NJ: Princeton University Press.

16. Canova, F., Sala, L. (2005). Back to square one: identification issues in DSGE models. Manuscript, IGIER and Bocconi University.

17. Chang, Y., Gomes, J., Schorfheide, F. (2002). Learning-by-doing as a propagation mechanism. American Economic Review 92 (5): 1498 – 1520.

18. Chang, Y., Schorfheide, F. (2003). Labor-supply shifts and economic fluctuations. Journal of Monetary Economics 50 (8): 1751 – 1768.

19. Chib, S., Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. American Statistician 49 (4): 327 – 335.

20. Chib, S., Jeliazkov, I. (2001). Marginal likelihood from the Metropolis-Hastings output. Journal of the American Statistical Association 96 (453): 270 – 281.

21. Christiano, L. J. (2002). Solving dynamic equilibrium models by a methods of undetermined coefficients. Computational Economics 20 (1–2): 21 – 55.

22. Christiano, L. J., Eichenbaum, M. (1992). Current real-business cycle theories and aggregate labor-market fluctuations. American Economic Review 82 (3): 430 – 450.

23. Christiano, L. J., Eichenbaum, M., Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. Journal of Political Economy 113 (1): 1 – 45.

24. Collard, F., Juillard, M. (2001). Accuracy of stochastic perturbation methods: the case of asset pricing models. Journal of Economic Dynamics and Control 25 (6–7): 979 – 999.

25. Crowder, M. (1988). Asymptotic expansions of posterior expectations, distributions, and densities for stochastic processes. Annals of the Institute for Statistical Mathematics 40:297–309.

26. de Walque, G., Wouters, R. (2004). An open economy DSGE model linking the Euro area and the U.S. economy. Manuscript, National Bank of Belgium.

27. DeJong, D. N., Ingram, B. F. (2001). The cyclical behavior of skill acquisition. Review of Economic Dynamics 4 (3): 536 – 561.

28. DeJong, D. N., Ingram, B. F., Whiteman, C. H. (1996). A Bayesian approach to calibration. Journal of Business and Economic Statistics 14 (1): 1 – 9.

29. DeJong, D. N., Ingram, B. F., Whiteman, C. H. (2000). A Bayesian approach to dynamic macroeconomics. Journal of Econometrics 98 (2): 203 – 223.

30. Del Negro, M. (2003). Fear of floating: a structural estimation of monetary policy in Mexico. Manuscript, Federal Reserve Bank of Atlanta.

31. Del Negro, M., Schorfheide, F. (2004). Priors from general equilibrium models for VARs. International Economic Review 45 (2): 643 – 673.

32. Del Negro, M., Schorfheide, F., Smets, F., Wouters, R. (2006). On the fit and forecasting performance of new Keynesian models. Journal of Business and Economic statistics, forthcoming.

33. Den Haan, W. J., Marcet, A. (1994). Accuracy in simulation. Review of Economic Studies 61 (1): 3 – 17.

34. Diebold, F. X., Ohanian, L. E., Berkowitz, J. (1998). Dynamic equilibrium economies: A framework for comparing models and data. Review of Economic Studies 65 (3): 433 – 452.

35. Dridi, R., Guay, A., Renault, E. (2007). Indirect inference and calibration of dynamic stochastic general equilibrium models. Journal of Econometrics 136 (2): 397 – 430.

36. DYNARE (2005). software available from http://cepremap.cnrs.fr/dynare/.

37. Fernández-Villaverde, J., Rubio-Ramírez, J. F. (2004). Comparing dynamic equilibrium models to data: A Bayesian approach. Journal of Econometrics 123 (1): 153 – 187.

38. Fernández-Villaverde, J., Rubio-Ramírez, J. F. (2005). Estimating dynamic equilibrium economies: linear versus non-linear likelihood. Journal of Applied Econometrics 20 (7): 891 – 910.

39. Galí, J., Rabanal, P. (2005). Technology shocks and aggregate fluctuations: how well does the RBC model fit postwar U.S. data. In: Gertler, M., Rogoff, K., eds. NBER Macroeconomics Annual 2004. Cambridge, MA: MIT Press, pp. 225 – 228.

40. GAUSS (2004). Software available from http://www.aptech.com/.

41. Gelfand, A. E., Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. Journal of the Royal Statistical Society Series B 56 (3): 501 – 514.

42. Gelman, A., Carlin, J., Stern, H., Rubin, D. (1995). Bayesian Data Analysis. New York: Chapman & Hall.

43. Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. Econometrica 57 (6): 1317 – 1339.

44. Geweke, J. (1999a). Computational experiments and reality. Manuscript, University of Iowa.

45. Geweke, J. (1999b). Using simulation methods for Bayesian econometric models: inference, development and communication. Econometric Reviews 18 (1): 1 – 126.

46. Geweke, J. (2005). Contemporary Bayesian Econometrics and Statistics. Hoboken: John Wiley & Sons.

47. Good, I. J. (1952). Rational decisions. Journal of the Royal Statistical Society Series B 14 (1): 107 – 114.

48. Gordon, N. J., Salmond, D. J., Smith, A. F. M. (1993). Novel approach to non-linear and non-gaussian Bayesian state estimation. IEE Proceedings-F 140 (2): 107 – 113.

49. Hammersley, J., Handscomb, D. (1964). Monte Carlo Methods. New York: John Wiley.

50. Hansen, L. P., Heckman, J. J. (1996). The empirical foundations of calibration. Journal of Economic Perspectives 10 (1): 87 – 104.

51. Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chain and their applications. Biometrika 57 (1): 97 – 109.

52. Ingram, B. F., Whiteman, C. M. (1994). Supplanting the Minnesota prior: forecasting macroeconomic time series using real business cycle model priors. Journal of Monetary Economics 34 (3): 497 – 510.

53. Ireland, P. N. (2004). A method for taking models to the data. Journal of Economic Dynamics and Control 28 (6): 1205 – 1226.

54. Jin, H.-H., Judd, K. L. (2002). Perturbation methods for general dynamic stochastic models. Manuscript, Hoover Institute.

55. Judd, K. L. (1998). Numerical Methods in Economics. Cambridge, MA: MIT Press.

56. Justiniano, A., Preston, B. (2004). Small open economy DSGE models – specification, estimation, and model fit. Manuscript, International Monetary Fund and Columbia University.

57. Kass, R. E., Rafterty, A. E. (1995). Bayes factors. Journal of the American Statistical Association 90 (430): 773 – 795.

58. Kim, J. (2000). Constructing and estimating a realistic optimizing model of monetary policy. Journal of Monetary Economics 45 (2): 329 – 359.

59. Kim, J., Kim, S. H. (2003). Spurious welfare reversal in international business cycle models. Journal of International Economics 60 (2): 471 – 500.

60. Kim, J., Kim, S. H., Schaumburg, E., Sims, C. A. (2005). Calculating and using second order accurate solutions of discrete time dynamic equilibirum models. Manuscript, Princeton University.

61. Kim, J.-Y. (1998). Large sample properties of posterior densities, bayesian information criterion, and the likelihood principle in non-stationary time series models. Econometrica 66 (2): 359 – 380.

62. Kim, K., Pagan, A. (1995). Econometric analysis of calibrated macroeconomic models. In: Pesaran, H., Wickens, M., eds. Handbook of Applied Econometrics: Macroeconomics. Oxford: Basil Blackwell, pp. 356 – 390.

63. Kim, S., Shephard, N., Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH Models. Review of Economic Studies 65 (3): 361 – 393.

64. King, R. G., Watson, M. W. (1998). The solution of singular linear difference systems under rational expectations. International Economic Review 39 (4): 1015-1026.

65. Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian non-linear state space models. Journal of Computational and Graphical Statistics 5 (1): 1 – 25.

66.   Klein, P. (2000). Using the generalized Schur form to solve a multivariate linear rational expectations model. Journal of Economic Dynamics and Control 24 (10): 1405-1423.

67.   Kydland, F. E., Prescott, E. C. (1982). Time to build and aggregate fluctuations. Econometrica 50 (6): 1345 – 1370.

68.   Kydland, F. E., Prescott, E. C. (1996). The computational experiment: an econometric tool. Journal of Economic Perspectives 10 (1): 69 – 85.

69.   Laforte, J. P. (2004). Comparing monetary policy rules in an estimated equilibrium model of the US economy. Manuscript, Federal Reserve Board of Governors.

70.   Lancaster, A. (2004). An Introduction to Modern Bayesian Econometrics. Oxford: Blackwell Publishing.

71.   Landon-Lane, J. (1998). Bayesian Comparison of Dynamic Macroeconomic Models. Ph.D. dissertation, University of Minnesota.

72.   Leeper, E. M., Sims, C. A.. (1994). Toward a modern macroeconomic model usable for policy analysis. In: Stanley, F., Rotemberg, J. J., eds. NBER MAcroeconomics Annual 1994. Cambridge, MA: MIT Press, pp. 81 – 118.

73.   Levin, A. T., Alexei O., Williams, J. C., Williams, N. (2006). Monetary policy under uncertainty in micro-founded macroeconometric models. In: Mark G., Rogoff, K., eds. NBER Macroeconomics Annual 2005. Cambridge, MA: MIT Press, pp. 229 – 287.

74.   Lubik, T. A., Schorfheide, F. (2003). Do central banks respond to exchange rate fluctuation – a structural investigation. Manuscript, University of Pennsylvania.

75.   Lubik, T. A., Schorfheide, F. (2004). Testing for indeterminacy: an application to US monetary policy. American Economic Review 94 (1): 190 – 217.

76.   Lubik, T. A., Schorfheide, F. (2006). A Bayesian look at new open economy macroeconomics. In: Mark, G., Rogoff, K., eds. NBER Macroeconomics Annual 2005. Cambridge, MA: MIT Press, pp. 313 – 366.

77.   MATLAB (2005). Version 7. Software available from www.mathworks.com/.

78.   McGrattan, E. R. (1994). The macroeconomic effects of distortionary taxation. Journal of Monetary Economics 33 (3): 573 – 601.

79.   Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E. (1953). Equation of state calculations by fast computing machines. Journal of Chemical Physics 21 (6): 1087 – 1092.

80.   Onatski, A., Williams, N. (2004). Empirical and policy performance of a forward-looking monetary model. Manuscript, Princeton University.

81. Otrok, C. (2001). On measuring the welfare cost of business cycles. Journal of Monetary Economics 47 (1): 61 – 92.

82. Phillips, P. C. B. (1996). Econometric model determination. Econometrica 64 (4): 763 – 812.

83. Pitt, M. K., Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. Journal of the American Statistical Association 94 (446): 590 – 599.

84. Poirier, D. (1998). Revising beliefs in nonidentified models. Econometric Theory 14(4):483–509.

85. Rabanal, P., Rubio-Ramírez, J. F . (2005a). Comparing new Keynesian models in the Euro area: A Bayesian approach. Federal Reserve Bank of Atlanta Working Paper 2005–8.

86. Rabanal, P., Rubio-Ramírez, J. F. (2005b). Comparing new Keynesian models of the business cycle: A Bayesian approach. Journal of Monetary Economics 52 (6): 1152 – 1166.

87. Rabanal, P. , Tuesta, V. (2006). Euro-Dollar real exchange rate dynamics in an estimated two-country model: What is important and what is not. CEPR Discussion Paper No. 5957.

88. Robert, C., Casella, G. (1999). Monte Carlo Statistical Methods. New York: Springer Verlag.

89. Rotemberg, J. J., Woodford, M. (1997). An optimization-based econometric framework for the evaluation of monetary policy. In: Bernanke, B. S., Rotemberg, J. J., eds. NBER Macroeconomics Annual 1997. Cambridge, MA: MIT Press, pp. 297 – 346.

90. Sargent, T. J. (1989). Two models of measurements and the investment accelerator. Journal of Political Economy 97 (2): 251 – 287.

91. Schmitt-Grohé, S., Uribe, M. (2006). Optimal simple and implementable monetary and fiscal rules. Journal of Monetray Economics. Duke University.

92. Schorfheide, F. (2000). Loss function-based evaluation of DSGE models. Journal of Applied Econometrics 15 (6): 645 – 670.

93. Schorfheide, F. (2005). Learning and monetary policy shifts. Review of Economic Dynamics 8 (2): 392 – 419.

94. Sims, C. A. (1996). Macroeconomics and methodology. Journal of Economic Perspectives 10 (1): 105 – 120.

95. Sims, C. A. (2002). Solving linear rational expectations models. Computational Economics 20 (1–2): 1 – 20.

96. Sims, C. A . (2003). Probability models for monetary policy decisions. Manuscript, Princeton University.

97. Smets, F., Wouters, R. (2003). An estimated dynamic stochastic general equilibrium model of the Euro area. Journal of European Economic Association 1 (5): 1123 – 1175.

98. Smets, F., Wouters, R. (2005). Comparing shocks and frictions in US and Euro area business cycles: A Bayesian DSGE Approach. Journal of Applied Econometrics 20 (2): 161 – 183.

99. Smith, A. A. (1993). Estimating non-linear time-series models using simulated vector autoregressions. Journal of Applied Econometrics 8: S63 – S84.

100. Swanson, E., Anderson, G., Levin, A. (2005). Higher-Order perturbation solutions to dynamic, discrete-time rational expectations models. Manuscript, Federal Reserve Bank of San Fransisco.

101. Taylor, J. B., Uhlig, H. (1990). Solving non-linear stochastic growth models: a comparison of alternative solution methods. Journal of Business and Economic Statistics 8 (1): 1 – 17.

102. Uhlig, H. (1999). A toolkit for analyzing non-linear dynamic stochastic models easily. In: Marimón, R., Scott, A., eds. Computational Methods for the Study of Dynamic Economies. Oxford, UK: Oxford University Press, pp. 30 – 61.

103. Walker, A. M. (1969). On the asymptotic behaviour of posterior distributions. Journal of the Royal Statistical Society Series B 31 (1): 80 – 88.

104. Watson, M. W. (1993). Measures of fit for calibrated models. Journal of Political Economy 101 (6): 1011 – 1041.

105. White, H. (1982). Maximum likelihood estimation of misspecified models. Econometrica 50 (1): 1 – 25.

106. Woodford, M. (2003). Interest and Prices: Foundations of a Theory of Monetary Policy. Princeton, NJ: Princeton University Press.

107. Zeller, A. (1971). Introduction to Bayesian Inference in Econometrics. New York: John Wiley & Sons.