DISCUSSION PAPER SERIES

No. 5207

BAYESIAN ANALYSIS OF DSGE MODELS

Sungbae An and Frank Schorfheide

INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP5207.asp**

ISSN 0265-8003

BAYESIAN ANALYSIS OF DSGE MODELS

Sungbae An, University of Pennsylvania
Frank Schorfheide, University of Pennsylvania and CEPR

Discussion Paper No. 5207
September 2005

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 5207

September 2005

ABSTRACT

Bayesian Analysis of DSGE Models*

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 a reference model, as well as the estimation of second-order accurate solutions of DSGE models. These methods are applied to data generated from a linearized DSGE model, a vector autoregression that violates the cross-coefficient restrictions implied by the linearized DSGE model, and a DSGE model that was solved with a second-order perturbation method.

JEL Classification: C11, C32, C51 and C52
Keywords: Bayesian analysis, DSGE models, model evaluation and vector autoregressions

Sungbae An
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104
USA

Email: sungbae@ssc.upenn.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=163125

Frank Schorfheide
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
USA

Email: schorf@ssc.upenn.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=158108

* Schorfheide gratefully acknowledges financial support from the Alfred P. Sloan Foundation.

Submitted 08 August 2005

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, Ingram, and Whiteman (1996), Diebold, Ohanian, and Berkowitz (1998), Geweke (1999b), Schorfheide (2000), Dridi, Guay, Renault (2004), and Bierens (2005). 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 useful tool for forecasting and quantitative policy analysis in macroeconomics. Moreover, due 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 build around a likelihood function derived from the DSGE model. The econometric analysis has to cope with several challenges, including potential model misspecification, identification problems, and the computational burden. 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

[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 (2005) which discusses both classical and Bayesian methods for the analysis of DSGE models.

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 model. 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 a small-scale New Keynesian DSGE model along the lines of Woodford (2003). 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 three data sets: one from a first-order accurate solution of the DSGE model presented in Section 2, one from the second-order accurate solution, and a third data set that is generated from a VAR that does not satisfy the cross-coefficient restrictions implied by the DSGE model.

Due to the computational burden associated with the construction of the likelihood function for nonlinear solutions of the DSGE model, most of the empirical literature estimated linearized DSGE models. Section 4 describes a Random Walk Metropolis Algorithm and an Importance Sampler that can be used to evaluate 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: model checking along the lines of Box (1980), model comparisons based on posterior odds, and comparisons of posterior distributions for important population characteristics such as impulse responses and autocovariances between the DSGE model and a reference model. We illustrate these techniques based on the datas generated from the DSGE model and the VAR. In Section 6 we construct posterior distributions for DSGE model parameters based on a second-order accurate solution. In the context of our simulation experiment we find that the posteriors obtained with the nonlinear techniques are more concentrated than the posteriors computed with the linear approximation. Finally, Section 7 concludes and provides an outlook on future work.

2 A Prototypical DSGE Model

The model economy consists of a final good producing firm, a continuum of intermediate goods producing firms, households, 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). We abstract from wage rigidities and capital accumulation. More elaborate versions of the model can be found in Smets and Wouters (2003) and Christiano, Eichenbaum, and Evans (2005).

2.1 The Agents and Their Decision Problems

The perfectly competitive, representative, final good 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 following 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

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

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

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

Here, Q_{t+s} is the marginal value of a unit of the consumption good 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. This assumption assures 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

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

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 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 households' budget constraint is of the form

$$P_tC_t+B_t+M_t-M_{t-1}+T_t=P_tW_tH_t+R_{t-1}B_{t-1}+P_tD_t \tag{8}$$

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

The monetary policy is described by an interest rate feedback rule of the form

$$R_t=R_t^{*\,1-\rho_R}R_{t-1}^{\rho_R}e^{\epsilon_{R,t}}, \tag{9}$$

where $\epsilon_{R,t}$ 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:

$$R_t^*=r\pi^*\left(\frac{\pi_t}{\pi^*}\right)^{\psi_1}\left(\frac{Y_t}{Y_t^*}\right)^{\psi_2}\quad\text{(output gap rule specification }\mathcal{M}_1) \tag{10}$$

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

$$R_t^*=r\pi^*\left(\frac{\pi_t}{\pi^*}\right)^{\psi_1}\left(\frac{Y_t}{\gamma Y_{t-1}}\right)^{\psi_2}\quad\text{(output growth rule specification }\mathcal{M}_2). \tag{11}$$

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

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

2.2 Exogenous Processes

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

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

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 σ_z, σ_g, σ_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

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

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

$$1 = \beta I\!E_t\left[\left(\frac{C_{t+1}/A_{t+1}}{C_t/A_t}\right)^{-\tau}\frac{A_t}{A_{t+1}}\frac{R_t}{\pi_{t+1}}\right] \tag{16}$$

$$\begin{aligned} 1 = {} & \frac{1}{\nu}\left[1-\left(\frac{C_t}{A_t}\right)^{\tau}\right]+\varphi(\pi_t-\pi)\left[\left(1-\frac{1}{2\nu}\right)\pi_t+\frac{\pi}{2\nu}\right] \\ & -\varphi\beta I\!E_t\left[\left(\frac{C_{t+1}/A_{t+1}}{C_t/A_t}\right)^{-\tau}\frac{Y_{t+1}/A_{t+1}}{Y_t/A_t}(\pi_{t+1}-\pi)\pi_{t+1}\right]. \end{aligned} \tag{17}$$

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

$$Y_t^* = (1-\nu)^{1/\tau}A_t g_t, \tag{18}$$

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

$$r = \frac{\gamma}{\beta}, \quad R = r\pi^*, \quad c = (1-\nu)^{1\tau}, \quad \text{and} \quad y = g(1-\nu)^{1\tau}. \tag{19}$$

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

$$1 = I\!E_t\left[e^{-\tau\hat{c}_{t+1}+\tau\hat{c}_t+\hat{R}_t-\hat{z}_{t+1}-\hat{\pi}_{t+1}}\right] \tag{20}$$

$$\begin{aligned} \frac{1-\nu}{\nu\varphi\pi^2}\left(e^{\tau\hat{c}_t}-1\right) = {} & \left(e^{\hat{\pi}_t}-1\right)\left[\left(1-\frac{1}{2\nu}\right)e^{\hat{\pi}_t}+\frac{1}{2\nu}\right] \\ & -\beta I\!E_t\left[\left(e^{\hat{\pi}_{t+1}}-1\right)e^{-\tau\hat{c}_{t+1}+\tau\hat{c}_t+\hat{y}_{t+1}-\hat{y}_t+\hat{\pi}_{t+1}}\right] \end{aligned} \tag{21}$$

$$e^{\hat{c}_t-\hat{y}_t} = e^{-\hat{g}_t}-\frac{\varphi\pi^2 g}{2}\left(e^{\hat{\pi}_t}-1\right)^2 \tag{22}$$

$$\hat{R}_t = \rho_R\hat{R}_{t-1}+(1-\rho_R)\psi_1\hat{\pi}_t+(1-\rho_R)\psi_2\left(\hat{y}_t-\hat{g}_t\right)+\epsilon_{R,t} \tag{23}$$

$$\hat{g}_t = \rho_g\hat{g}_{t-1}+\epsilon_{g,t} \tag{24}$$

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

For the output growth rule specification, Equation (23) is replaced by

$$\hat{R}_t = \rho_R\hat{R}_{t-1}+(1-\rho_R)\psi_1\hat{\pi}_t+(1-\rho_R)\psi_2\left(\Delta\hat{y}_t+\hat{z}_t\right)+\epsilon_{R,t}. \tag{26}$$

2.4 Model Solutions

Equations (20) to (25) form a nonlinear rational expectations system in the variables $\hat{y}_t$, $\hat{c}_t$, $\hat{\pi}_t$, $\hat{R}_t$, $\hat{g}_t$, 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]

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

The solution of the rational expectations system takes the form

$$s_t = \Phi(s_{t-1}, \epsilon_t; \theta). \tag{27}$$

From an econometric perspective, s_t can be viewed as a (partially latent) state vector in a nonlinear state space model and (27) 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. A linearization of Equations (20) to (22) yields

$$\hat{c}_t = \hat{c}_{t+1} - \frac{1}{\tau}[\hat{R}_t - \hat{\pi}_{t+1} - \hat{z}_{t+1}] \tag{28}$$

$$\hat{\pi}_t = \beta \mathbb{E}_t[\hat{\pi}_{t+1}] + \kappa \hat{c}_t \tag{29}$$

$$\hat{y}_t = \hat{c}_t + \hat{g}_t, \tag{30}$$

where

$$\kappa = \tau \frac{1-\nu}{\nu \pi^2 \varphi}. \tag{31}$$

Equations (28) to (30) combined with (23) to (25) 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), Klein (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 j'th element of s_t takes the form

$$s_{j,t} = \sum_{i=1}^{J} \Phi_{j,i}^{(s)} s_{i,t-1} + \sum_{l=1}^{n} \Phi_{j,l}^{(\epsilon)} \epsilon_{l,t}, \quad j = 1, \ldots, J. \tag{32}$$

[2] In the output growth specification s_t also contains $\hat{y}_{t-1}$.

Here the coefficients $\Phi_{j,i}^{(s)}$ and $\Phi_{j,l}^{(\epsilon)}$ 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 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 (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 follows

$$U(\hat{c}) = U(0) + U^{(1)}(0)\hat{c} + \frac{1}{2}U^{(2)}(0)\hat{c}^2 + o(|\hat{c}|^2), \tag{33}$$

where $\hat{c}$ is the percentage deviation of consumption from its steady state and the superscript (i) denotes the i'th 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

$$\hat{c} = C^{(1)}(c)\sigma\epsilon + \frac{1}{2}C^{(2)}(c)\sigma^2\epsilon^2 + o_p(\sigma^2). \tag{34}$$

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 (23). A second order approximation of both U and C leads to

$$I\!E[U(\hat{c})] = U(0) + \frac{1}{2}U^{(2)}(0)C^{(1)^2}(c)\sigma^2 + \frac{1}{2}U^{(1)}(0)C^{(2)}(c)\sigma^2 + o(\sigma^2). \tag{35}$$

Using a second order expansion of the utility function together with a first-order expansion of consumption ignores the second term in (35) and is only appropriate if either $U^{(1)}(0)$ or $C^{(2)}(c)$ are negligible.

A second-order accurate solution to the DSGE model can be obtained from a second-order expansion of the equilibrium conditions (20) to (25). Algorithms to construct such solutions have been developed by Judd (1998), Collard and Juillard (2001), Jin and Judd (2002), Schmitt-Grohe and Uribe (2004), Kim, Kim, Schaumburg and Sims (2005), and Swanson, Anderson, and Levin (2005). The resulting state transition equation can be expressed as

$$\begin{aligned} s_{j,t} \;=\; & \Phi_j^{(0)} + \sum_{i=1}^{J}\Phi_{j,i}^{(s)}s_{i,t-1} + \sum_{l=1}^{n}\Phi_{j,l}^{(\epsilon)}\epsilon_{l,t} + \sum_{i=1}^{J}\sum_{l=1}^{J}\Phi_{j,il}^{(ss)}s_{i,t-1}s_{l,t-1} \\ & + \sum_{i=1}^{J}\sum_{l=1}^{n}\Phi_{j,il}^{(s\epsilon)}s_{i,t-1}\epsilon_{l,t} + \sum_{i=1}^{n}\sum_{l=1}^{n}\Phi_{j,il}^{(\epsilon\epsilon)}\epsilon_{i,t}\epsilon_{l,t}. \end{aligned} \tag{36}$$

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-Grohe and Uribe's (2004) algorithm to obtain a second-order accurate solution.

Noting that a second-order approximation is an extension of a first-order approximation, i.e., linear terms in (32) and (36) match exactly, a higher-order perturbation is readily available and based on the same method with higher-order Taylor expansion. While perturbation methods approximate policy functions 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, Fernández-Villaverde, and Rubio-Ramírez (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 we have observations on quarter-to-quarter per capita GDP growth rates (YGR), annualized quarter-to-quarter inflation rates (INFL), and annualized nominal interest rates (INT) in percentages:

$$\begin{aligned} \ln(YGR_t) &= \gamma^Q + \hat{y}_t - \hat{y}_{t-1} + \hat{z}_t \\ \ln(INFL_t) &= \pi^{(A)} + 4 \times \hat{\pi}_t \\ \ln(INT_t) &= \pi^{(A)} + r^{(A)} + 4 \times \gamma^{(Q)} + 4 \times \hat{R}_t. \end{aligned} \tag{37}$$

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

$$\gamma = 1 + \frac{\gamma^{(Q)}}{100}, \quad \beta = \frac{1}{1 + r^{(A)}/400}, \quad \pi = 1 + \frac{pi^{(A)}}{400}.$$

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 (31). 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 (φ, ν) or alternatively by (κ, ν). 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) and Christiano, Eichenbaum and Evans (2005), 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 (20), the price setting equation of the intermediate goods producing firms (21), or the monetary policy rule (23), 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, possible lack of identification of parameters of interest, computational burden. We will subsequently elaborate on 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 estimated the DSGE model. In the context of the model developed in Section 2 y_t is composed of $\ln(YGR_t)$, $\ln(INFL_t)$, $\ln(INT_t)$. 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 $\mathcal{L}(\theta|Y)$ and the posterior density by $p(\theta|Y)$.

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 equals the number of observables to which the model is fitted.

A second source of misspecification that is more difficult to correct are 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, Schorfheide, Smets, and Wouters (2004) 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 out-of-sample information. 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 micro-level evidence. However, due to the stylized nature and the resulting misspecification of most DSGE model, 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 re-weighted by a prior density. The prior can bring to bear information that is not contained in the estimation sample Y. Unlike in a maximum likelihood approach that uses extraneous information to fix elements of the parameter vector θ, the prior density allows to weight information about different parameters according to its reliability. 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 out-of-sample information that has been used to construct the prior distribution, then the marginal data density, defined as

$$p(Y) = \int \mathcal{L}(\theta|Y)p(\theta)d\theta \tag{38}$$

will be low. Hence, in posterior odds comparisons 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 due 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 (2005). 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:

$$\mathcal{M}_1: \quad y_t = \frac{1}{\alpha}\mathbb{E}_t[y_{t+1}] + u_t, \quad u_t = \rho u_{t-1} + \epsilon_t, \quad \epsilon_t \sim iid\Big(0, (1-\rho\alpha)^2\Big). \tag{39}$$

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:

$$\mathcal{M}_2: \quad y_t = \frac{1}{\alpha}\mathbb{E}_t[y_{t+1}] + \phi y_{t-1} + u_t, \quad u_t = \epsilon_t, \quad \epsilon_t \sim iid\left(0, \left[\frac{\alpha + \sqrt{\alpha^2 - 4\phi\alpha}}{2\alpha}\right]^2\right). \tag{40}$$

For both specifications, the law of motion of y_t is

$$y_t = \psi y_{t-1} + \eta_t, \quad \eta_t \sim iid(0,1). \tag{41}$$

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:

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

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 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 $\mathcal{L}(\rho, \alpha|Y) = \bar{\mathcal{L}}(\rho)$. Straightforward manipulations of Bayes Theorem yield

$$p(\rho, \alpha|Y) = p(\rho|Y)p(\alpha|\rho). \tag{42}$$

[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 $\mathcal{M}_2$.

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 directly detect 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 nonlinear 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.

3.3 Computational Burden

Numerical methods will be used to evaluate posterior moments. To generate a parameter draw from the posterior distribution the rational expectations model, in our case comprised of (20) to (25), has to be solved and the likelihood function has to be evaluated. The computational burden related to solving the model depends on its size and the desired accuracy. Linear approximation methods are based on generalized eigenvalue decompositions of the system matrices associated with the rational expectation system. Quadratic approximations are obtained by first constructing a linear approximation and then solving a discrete Lyapunov equation to obtain the second-order terms for the model solution. If the model is solved by a linear approximation method then the likelihood function can be evaluated with the Kalman filter. If on the other hand, the model is solved with a nonlinear method such as a quadratic approximation, then a nonlinear particle filter is necessary to evaluate the likelihood function and the computational burden is significantly larger. Based on a sample of 80 observations we can generate 1,000 draws with 40,000 particles in 1 hour and 40 minutes (6.0 seconds per draw) on AMD64 3000+ machine with 1GB RAM. Linear methods are more than 200 times faster: 1,000 draws are generated in 24 seconds in our Matlab routines and 3 seconds in our GAUSS routines.

The model considered in this paper is small compared to the more elaborate model of Smets and Wouters (2003) or the two-country models in Lubik and Schorfheide (2005) or de Walque and Wouters (2004). While the time it takes to generate a single draw for the

Markov chain of the posterior simulator is increasing in model dimension and sample size, the speed of convergence of the Markov chain to its ergodic distribution is a function of the dimensionality of the parameter space. While the model examined in this paper has only 13 (15) parameters, the Smets and Wouters (2003) model has 33 parameters. The larger the parameter space the more likely it is that the MCMC algorithm explores the posterior distribution only locally in the neighborhood of the mode. In Section 4 we will provide an example in which the posterior has at least two modes that are separated by a deep valley. In this example the widely used random walk Metropolis algorithm only provides local approximations in the neighborhood of the two modes. This problem is likely to be amplified in the large scale DSGE models that are currently being developed by many central banks.

3.4 Priors

As indicated in our previous discussion, prior distributions will play an important role in the estimation of DSGE models. They might down-weigh 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 (2005) 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 micro-evidence 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

[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.

for various sample moments and make sure that the prior does not place to much mass on values that are clearly at odds with the pre-sample information or with beliefs about the relative importance of the various structural shocks.

Table 1 lists the marginal prior distributions for the structural parameters of the DSGE model. These priors are adopted from Lubik and Schorfheide (2005). For convenience, it is typically 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. The distribution specified in Table 1 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.

3.5 The Road Ahead

Throughout this paper we are estimating the models $\mathcal{M}_1$ and $\mathcal{M}_2$ based on simulated data to illustrate the methods that have been used in the literature to estimate and evaluate DSGE models and study their performance. The sample size is 80 observations which is a fairly realistic number for the estimation of monetary DSGE model. Many industrialized countries experienced a high inflation episode in the 1970s which 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. We consider four data sets: Data Sets 1-$\mathcal{M}_1$ and 1-$\mathcal{M}_2$ have been generated directly from the first-order accurate solution of the DSGE models $\mathcal{M}_1$ and $\mathcal{M}_2$. Data Set 2 has been generated from a VAR whose coefficients do not exactly satisfy the restrictions implied by the DSGE model. Thus, the DSGE model is misspecified relative to the data generating process. Details on how Data Set 2 is generated will be provided in Section 5. Data Set 3 is generated from the quadratic approximation of $\mathcal{M}_1$. The last two columns of Table 1 contain the parameter values that were used to generate Data Set 1-$\mathcal{M}_i$ and 3. The parameters used for Data Set 1 resemble empirical estimates obtained from post-1982 U.S. data. The parameters for Data Set 3 have been

adjusted so that the first two moments of output growth, inflation, and interest rates mirror the moments in Data Set 1.

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 $\mathcal{M}_1$ and $\mathcal{M}_2$.

4.1 Posterior Computations

We consider a Random-Walk Metropolis (RWM) Algorithm and an Importance Sampling (IS) Algorithm to generate draws from the posterior distribution of θ. Both algorithms require the evaluation of $\mathcal{L}(\theta|Y)p(\theta)$. 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 (32). If the parameter value θ implies indeterminacy (or non-existence of a stable rational expectations solution) then $\mathcal{L}(\theta|Y)p(\theta)$ is set to zero. If a unique stable solution exists then the Kalman filter[5] is used to evaluation the likelihood function associated with the linear state-space system (32) and (37). 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 comprised 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 used to generated draws from the posterior distribution of DSGE model parameters by Schorfheide (2000) and Otrok (2001). We will use the following implementation based on Schorfheide (2000):[6]

[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 4 observations in the sample: $\mathcal{L}(\theta|Y)/\mathcal{L}(\theta|y_1,\ldots,y_4)$. These 4 observations are later used to initialize the lags of a VAR.

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

Random-Walk Metropolis (RWM) Algorithm

1. Use a numerical optimization routine to maximize $\ln \mathcal{L}(\theta|Y) + \ln p(\theta)$. 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 $\mathcal{N}(\tilde{\theta}^{(0)}, c_0^2\tilde{\Sigma})$.

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

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 algorithm generates a sequence of dependent draws from the posterior distribution of θ that can be averaged to approximate posterior moments. Geweke (1999a) 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 $I\!E[h(\theta)|Y]$.

Dejong, Ingram, and Whiteman (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
$$I\!E[h(\theta)|Y] = \int h(\theta)p(\theta|Y)d\theta = \int \frac{h(\theta)p(\theta|Y)}{q(\theta)}q(\theta)d\theta.$$
Draws from the posterior density $p(\theta|Y)$ are replaced by draws from the density $q(\theta)$ and re-weighted by the importance ratio $p(\theta|Y)/q(\theta)$ to obtain a numerical approximation of the posterior moment of interest. Hammersley and Handscomb (1964) were among the firsts 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 $\ln \mathcal{L}(\theta|Y) + \ln p(\theta)$. 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. Let $q(\theta)$ be the density of a multivariate t-distribution with mean $\tilde{\theta}$, covariance 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 $\tilde{w}_s = \mathcal{L}(\theta^{(s)}|Y)p(\theta^{(s)})/q(\theta^{(s)})$ 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 assigns small values to regions of the parameter space in which the posterior density is large, then most of the w_s's will be zero and the approximation method becomes 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 $\mathcal{M}_1$

We begin by estimating the output gap rule specification $\mathcal{M}_1$ based on Data Set 1-$\mathcal{M}_1$. We use the RWM Algorithm ($c_0 = 1$, $c = 0.3$) to generate 1,000,000 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 2-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 (29). 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 4 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, 0.1), (3.0, 2.0), (0.4, 0.1), and (0.6, 2.0), respectively. As before, we set $c = 0.3$, generate 1,000,000 draws for each chain, and plot every 5,000'th 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 4 chains. Despite different initializations, the means converge in the long-run.

We generate another set of 1,000,000 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 (1999a) and for the Metropolis chain we use Newey-West standard error estimates, truncated at 150 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, say, the difference between the posterior mode and the estimated posterior means.

4.3 Simulation Results for $\mathcal{M}_2$

We proceed by estimating the output growth rule version $\mathcal{M}_2$ of the DSGE model based on 80 observations generated from this model (Data Set 1-$\mathcal{M}_2$). Unlike the posterior for the

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

output gap version, the $\mathcal{M}_2$ posterior has (at least) two modes. One of the modes, which we label the 'high' mode, is located at $\tau = 2.06$ and $\psi_2 = 0.97$. The value of $\ln[\mathcal{L}(\theta|Y)p(\theta)]$ is equal to -175.48. The second ('low') mode is located at $\tau = 1.44$ and $\psi_2 = 0.81$ 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. 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 1,000,000 draws each for model $\mathcal{M}_2$ from 4 Markov chains that were initialized at the following values for $(\tau;\ \psi_2)$: (3.0; 0.1), (3.0; 2.0), (0.4; 0.1), and (0.6; 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 from the 4 Markov chains in panels (2,1) to (3,2). 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 4 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.

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

[8] The covariance matrices of the proposal distributions are obtained from the Hessians associated with the respective modes.

still provide a good approximation to the overall posterior distribution. We will revisit this point in Section 5.2.

4.4 Parameter Transformations

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, 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 comprised of (32) and (37). We will focus on variance decompositions in the remainder of this subsection.

Fluctuations of output growth, inflation, and nominal interest rate in model $\mathcal{M}_1$ are due to 3 shocks: technology growth shocks, government spending shocks, and monetary policy shocks. Hence, variance decompositions of the endogenous variables lie in a 3-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.

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) time-to-build model, and McGrattan (1994) 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, Ingram, and Whiteman (1996), and Geweke (1999b) 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, Ingram, and Whiteman (2000), Schorfheide (2000), and Otrok (2001). DeJong, Ingram, and Whiteman (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, Gomes, and Schorfheide (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-Ramırez (2004a) 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-Ramırez (2003, 2005) 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 (2004) 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, Eichenbaum, and Evans (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, Onatski, Williams, and Williams (2005) 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 (SOE) 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. Adolfsen, Laséen, Lindé, and Villani (2004) analyze a large-scale SOE model, that includes capital accumulation as well as numerous real and nominal frictions. Lubik and Schorfheide (2005), Rabanal and Tuesta (2005), and de Walque and Wouters (2004) have estimated multi-country 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 sampling based 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 model $\mathcal{M}$ is considered as discredited by the

data if one observes an event that is deemed very unlikely by $\mathcal{M}$. 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 Gelman, Carlin, Stern, and Rubin (1995).

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 sampling distribution of Y^{rep} given the current state of knowledge:

$$p(Y^{rep}|Y) = \int p(Y^{rep}|\theta)p(\theta|Y)d\theta. \tag{43}$$

Let $T(Y)$ 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

$$\int \{T(Y^{rep}) \geq T(Y)\}p(Y^{rep}|Y)dY^{rep}, \tag{44}$$

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

To illustrate the model checking idea we use a graphical approach. We consider the following data transformations or test statistics: the correlation between interest rates and inflation, the correlation between output growth and inflation, the standard deviations and the first order autocorrelations of inflation and interest rates, the correlation between inflation and lagged interest rates, and the correlation between inflation and lagged output growth. To obtain draws from the posterior predictive distribution of $T(Y^{rep})$ 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 $T(Y^{rep})$. The left panels of Figure 8 depict the resulting 200 draws from the posterior predictive distribution in bivariate scatter plots. Moreover, the intersection of the solid lines indicate the actual values $T(Y)$. Since Data Set 1 was generated from model $\mathcal{M}_1$ we would expect the estimated model to pass the posterior predictive check. Indeed, a visual inspection of the plots suggests that the actual values of the test statistics do not lie in regions that essentially

[9]To obtain draws from the unconditional distribution of y_t we initialize $s_0 = 0$ and generate 180 observations, discarding the first 100.

have zero probability under the predictive distribution generated by the estimated DSGE model.

We repeat the exercise for Data Set 2, which has not been generated from the DSGE model $\mathcal{M}_1$. The draws from the posterior predictive distribution together with the actual sample moments are presented in the right panels of Figure 8. Panel (3,2) provides some evidence of misspecification as the observed correlations between inflation and lagged interest and output growth rates appears to lie far in the tail of the posterior predictive distribution. Ultimately, the goal of a predictive check is to learn in which dimensions the model is at odds with the data. Our simulation exercise suggests that the DSGE model generates a correlation between inflation and lagged output growth that is too small.

5.2 Posterior Odds Comparisons

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 Raftery (1995). For concreteness, suppose in addition to the specifications $\mathcal{M}_1$ and $\mathcal{M}_2$ we consider a variant of $\mathcal{M}_1$, denoted by $\mathcal{M}_3$ in which prices are nearly flexible, that is, $\kappa = 5$. Moreover, according model $\mathcal{M}_4$ the central bank does not respond to output at all. If we are willing to place prior probabilities $\pi_{i,0}$ on the competing specifications then posterior model probabilities can be computed as follows:

$$\pi_{i,T} = \frac{\pi_{i,0} p(Y|\mathcal{M}_i)}{\sum_{j=1}^{4} \pi_{j,0} p(Y|\mathcal{M}_j)}, \quad j = 1, \ldots, 4. \tag{45}$$

The key object in the calculation of posterior probabilities is the the marginal data density $p(Y|\mathcal{M}_i)$, which is was defined in (38).

Posterior model probabilities can be used to weigh predictions from different models, or to justify the selection of a particular model specification and to condition all subsequent analysis on the selected model. It is straightforward to verify that under a 0-1 loss function, that is 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-Ramırez (2004a), 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

$$\ln p(Y|\mathcal{M}) = \sum_{t=1}^{T} \ln p(y_t|Y^{t-1},\mathcal{M}) = \sum_{t=1}^{T} \ln \left[\int p(y_t|Y^{t-1},\theta,\mathcal{M}) p(\theta|Y^{t-1},\mathcal{M}) d\theta \right], \quad (46)$$

$\ln p(Y|\mathcal{M})$ can be interpreted as 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 (1999a) modified harmonic mean estimator and Chib and Jeliazkov's (2001) algorithm.

Harmonic mean estimators are based on the following identity

$$\frac{1}{p(Y)} = \int \frac{f(\theta)}{\mathcal{L}(\theta|Y)p(\theta)} p(\theta|Y) d\theta, \quad (47)$$

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

$$\hat{p}_G(Y) = \left[\frac{1}{n_{sim}} \sum_{s=1}^{n_{sim}} \frac{f(\theta^{(s)})}{\mathcal{L}(\theta^{(s)}|Y)p(\theta^{(s)})} \right]^{-1}, \quad (48)$$

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

$$\begin{aligned} f(\theta) &= \tau^{-1}(2\pi)^{-d/2}|V_\theta|^{-1/2} \exp\left[-0.5(\theta-\bar{\theta})'V_\theta^{-1}(\theta-\bar{\theta})\right] \\ &\quad \times \left\{(\theta-\bar{\theta})'V_\theta^{-1}(\theta-\bar{\theta}) \leq F^{-1}_{\chi^2_d}(\tau)\right\}. \end{aligned} \quad (49)$$

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, $F_{\chi^2_d}$ 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 (48) are approximately constant.

Chib and Jeliazkov (2001) use the following equality as the basis for their estimator of the marginal data density

$$p(Y) = \frac{\mathcal{L}(\theta|Y)p(\theta)}{p(\theta|Y)}. \quad (50)$$

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

Thus,

$$\hat{p}_{CS}(Y) = \frac{\mathcal{L}(\tilde{\theta}|Y)p(\tilde{\theta})}{\hat{p}(\tilde{\theta}|Y)}, \tag{51}$$

where we replaced the generic θ in (50) by the posterior mode $\tilde{\theta}$. Within the RWM Algorithm denote the probability of moving from θ to ϑ by

$$\alpha(\theta, \vartheta|Y) = \min\ \{1, r(\theta, \vartheta|Y)\}, \tag{52}$$

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

$$\hat{p}(\tilde{\theta}|Y) = \frac{\frac{1}{n_{sim}} \sum_{s=1}^{n_{sim}} \alpha(\theta^{(s)}, \tilde{\theta}|Y) q(\theta^{(s)}, \tilde{\theta}|Y)}{J^{-1} \sum_{j=1}^{J} \alpha(\tilde{\theta}, \theta^{(j)}|Y)}, \tag{53}$$

where $\{\theta^{(s)}\}$ are sampled draws from the posterior distribution with the RWM Algorithm and $\{\theta^{(j)}\}$ are draws from $q(\tilde{\theta}, \theta|Y)$ given the fixed posterior mode value $\tilde{\theta}$.

We first use Geweke's harmonic mean estimator to approximate the marginal data densities associated with $\mathcal{M}_1$ and $\mathcal{M}_2$. The results are summarized in Figure 9. We calculate marginal data densities recursively (as a function of the number of MCMC draws) for the 4 chains that were initialized at different starting values as discussed in Section 4. For the output gap rule all 4 chains lead to estimates of around -196.7. The second panel of Figure 9 depicts the results for the output growth rule specification. As for the posterior means, the marginal data density estimates converge to two different limit points. The chains that were initialized near the low mode yield an estimate of about -206.9, whereas the other two chains generate an estimate of -197.4. If the valley that separates the two modes of the posterior divides the parameter space into two regions of equal prior probability, then according to the marginal data density estimates more than 99.9% of the posterior mass concentrates near the high mode. Unfortunately, it is difficult to compute the prior probability masses of the two regions and we do not attempt to do so in this paper. 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.

Based on Data Set 1-$\mathcal{M}_1$ we also estimate specification $\mathcal{M}_3$ in which prices are nearly flexible ($\kappa = 5$) and specification $\mathcal{M}_4$ in which the central bank does not respond to output ($\psi_2 = 0$). The marginal data density associated with $\mathcal{M}_3$ is -245.6, which translates into

a Bayes factor (ratio of posterior odds to prior odds) of approximately e^{49} in favor of $\mathcal{M}_1$. Hence, Data Set 1 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 $\mathcal{M}_4$ is equal to -201.9 and the Bayes factor of model $\mathcal{M}_1$ versus model $\mathcal{M}_4$ is only e^4. 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.

5.3 Comparisons to a Reference Model

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 vector autoregression (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 such a phenomenon is often called Lindley's paradox.

If the posterior odds favor the VAR over the DSGE model then it becomes important to further investigate the deficiencies of the structural model. This can be achieved by comparing the posterior distributions of interesting population characteristics such as autocovariances and impulse response functions obtained from the DSGE model and the VAR. The remainder of this subsection will focus on the construction of a vector autoregressive reference model that can be used for the evaluation of a DSGE model. We will refer to the benchmark model as DSGE-VAR, which has been proposed by Del Negro and Schorfheide (2004) and used to evaluated a large-scale New Keynesian model in Del Negro, Schorfheide, Smets, and Wouters (2004).

To construct a DSGE-VAR we proceed as follows. Consider a vector autoregressive

specification of the form

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + u_t, \quad I\!E[u_t u_t'] = \Sigma. \tag{54}$$

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 $I\!E_\theta^D[\cdot]$ be the expectation under DSGE model and define the autocovariance matrices

$$\Gamma_{XX}(\theta) = I\!E_\theta^D[x_t x_t'], \quad \Gamma_{XY}(\theta) = I\!E_\theta^D[x_t y_t'].$$

A VAR approximation of the DSGE model can be defined through the following restriction functions that relate the DSGE model parameters to the VAR parameters

$$\Phi^*(\theta) = \Gamma_{XX}^{-1}(\theta)\Gamma_{XY}(\theta), \quad \Sigma^*(\theta) = \Gamma_{YY}(\theta) - \Gamma_{YX}(\theta)\Gamma_{XX}^{-1}(\theta)\Gamma_{XY}(\theta). \tag{55}$$

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 4 lags are sufficient to generate a fairly precise approximation of model $\mathcal{M}_1$.[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 (54) with the following coefficient matrices:

$$\Phi = \Phi^*(\theta) + \Phi^\Delta, \quad \Sigma = \Sigma^*(\theta) + \Sigma^\Delta. \tag{56}$$

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

$$\begin{aligned} \Sigma|\theta &\sim \mathcal{IW}\Big(\lambda T\Sigma^*(\theta), \lambda T - k, n\Big) \\ \Phi|\Sigma,\theta &\sim \mathcal{N}\left(\Phi^*(\theta), \frac{1}{\lambda T}\Big[\Sigma^{-1} \otimes \Gamma_{XX}(\theta)\Big]^{-1}\right), \end{aligned} \tag{57}$$

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

[10] In principle one could start from a VARMA model to avoid the approximation error. However, the posterior computations would become 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.

model with parameters θ, see Del Negro and Schorfheide (2004). It has the property that the prior 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, Schorfheide, Smets, and Wouters (2004). 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 $\lambda \longrightarrow \infty$. 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 follows:

$$p_\lambda(\Phi, \Sigma, \theta|Y) = p_\lambda(\Phi, \Sigma|Y, \theta)p_\lambda(\theta|Y). \tag{58}$$

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

$$\begin{aligned} \Sigma|Y,\theta &\sim \mathcal{IW}\Big((1+\lambda)T\hat{\Sigma}_b(\theta), (1+\lambda)T-k, n\Big) \\ \Phi|Y,\Sigma,\theta &\sim \mathcal{N}\Big(\hat{\Phi}_b(\theta), \Sigma\otimes(\lambda T\Gamma_{XX}(\theta)+X'X)^{-1}\Big), \end{aligned} \tag{59}$$

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

$$\begin{aligned} \hat{\Phi}_b(\theta) &= (\lambda T\Gamma_{XX}(\theta)+X'X)^{-1}(\lambda T\Gamma_{XY}+X'Y) \\ &= \left(\frac{\lambda}{1+\lambda}\Gamma_{XX}(\theta)+\frac{1}{1+\lambda}\frac{X'X}{T}\right)^{-1}\left(\frac{\lambda}{1+\lambda}\Gamma_{XY}+\frac{1}{1+\lambda}\frac{X'Y}{T}\right) \\ \hat{\Sigma}_b(\theta) &= \frac{1}{(1+\lambda)T}\Big[(\lambda T\Gamma_{YY}(\theta)+Y'Y)-(\lambda T\Gamma_{YX}(\theta)+Y'X) \\ &\quad \times(\lambda T\Gamma_{XX}(\theta)+X'X)^{-1}(\lambda T\Gamma_{XY}(\theta)+X'Y)\Big]. \end{aligned}$$

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

$$\begin{aligned} p_\lambda(Y|\theta) &= \frac{|\lambda T\Gamma_{XX}(\theta)+X'X|^{-\frac{n}{2}}|(1+\lambda)T\hat{\Sigma}_b(\theta)|^{-\frac{(1+\lambda)T-k}{2}}}{|\lambda T\Gamma_{XX}(\theta)|^{-\frac{n}{2}}|\lambda T\Sigma^*(\theta)|^{-\frac{\lambda T-k}{2}}} \\ &\quad \times\frac{(2\pi)^{-nT/2}2^{\frac{n((1+\lambda)T-k)}{2}}\prod_{i=1}^n\Gamma[((1+\lambda)T-k+1-i)/2]}{2^{\frac{n(\lambda T-k)}{2}}\prod_{i=1}^n\Gamma[(\lambda T-k+1-i)/2]}. \end{aligned}$$

A derivation is provided in Del Negro and Schorfheide (2004).

A comparison of VAR and DSGE model impulse responses requires the identification of structural shocks in the context of the VAR. So far, the VAR given in (54) 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

$$u_t = \Sigma_{tr}\Omega\epsilon_t. \tag{60}$$

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 (54). 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

$$\left(\frac{\partial y_t}{\partial \epsilon_t'}\right)_{DSGE} = A_0(\theta) = \Sigma_{tr}^*(\theta)\Omega^*(\theta), \tag{61}$$

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

$$\left(\frac{\partial y_t}{\partial \epsilon_t'}\right)_{VAR} = \Sigma_{tr}\Omega. \tag{62}$$

To identify the DSGE-VAR, we maintain the triangularization of its covariance matrix Σ and replace the rotation Ω in (62) with the function $\Omega^*(\theta)$ that appears in (61). The rotation matrix is chosen such 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 $p_\lambda(\theta|Y)$.

2. Use Geweke's modified harmonic mean estimator to obtain a numerical approximation of $\hat{p}_\lambda(Y)$.

3. For each draw $\theta^{(s)}$ generate a pair $\Phi^{(s)}$, $\Sigma^{(s)}$, by sampling from the $\mathcal{IW}-\mathcal{N}$ distribution. Moreover, compute the orthonormal matrix $\Omega^{(s)}$ as described above.

All the results reported subsequently are based on a DSGE-VAR with $p = 4$ lags. Before we begin the numerical illustration we generate a second data set, Data Set 2, from the DSGE-VAR as follows. We use specification $\mathcal{M}_1$ in combination with the parameter values of θ listed in column 'DGP1' of Table 1. We set $T = 80$ and $\lambda = 0.5$ and draw VAR coefficient matrices Φ_0 and Σ_0 from their prior distribution (57). Conditional on Φ_0 and Σ_0 we simulate a sample of 80 observations from the VAR.

The first step of our analysis is to construct a posterior distribution for the parameter λ. We assume that λ lies on the grid $\Lambda = \{0.25, 0.5, 0.75, 1, 5, \infty\}$. 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 2 reports log marginal data densities for the DSGE-VAR as a function of λ for Data Set 1 and 2. If the VAR approximation of the DSGE model is accurate, then the values in the first and second row of Table 2 should coincide. 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(∞). For Data Set 1, which has been directly generated from the state space representation of $\mathcal{M}_1$, 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. With respect to Data Set 2 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, Schorfheide, Smets, and Wouters (2004). The value $\lambda = 0.75$ has the highest likelihood. The marginal data densities drop substantially for $\lambda > 1$ and $\lambda < 0.5$.

In Figures 11 and 12 we compare impulse response functions computed from the state space representation of the DSGE model and the DSGE-VAR($\hat{\lambda}$), where $\hat{\lambda}$ maximizes the marginal data densities. Under Data Set 1 $\hat{\lambda} = \infty$ and the posterior mean DSGE-VAR responses to a monetary policy shock and a technology growth shock are virtually indistinguishable from the DSGE model responses. The government spending shock, however, has only a transitory effect on output in the DSGE model, whereas it has a permanent effect in the estimated DSGE-VAR. This discrepancy is due to the VAR approximation error. More interesting is Figure 12 which is based on Data Set 2. Here $\hat{\lambda} = 0.75$ which allows for sizable discrepancy between DSGE-VAR and DSGE model dynamics. Notice that the identification scheme outlined above still delivers interpretable impulse response functions. While the contractionary monetary policy shock causes a rise in interest rates and a fall of inflation,

the response of output, in particular in the long-run, is associated with large uncertainty. According to the DSGE-VAR the long-run effect of a technology shock on output relative to inflation is larger than for the DSGE model. Moreover, a government spending shock seems to cause a drop in interest rates that cannot be reproduced by the DSGE model. Thus, in empirical applications the discrepancies between DSGE-VAR and DSGE model impulse responses can be used to gain insights about model deficiencies.

Finally, Figure 13 depicts scatter plots with draws from the posterior distributions of various population correlations and standard deviations based on Data Set 2. The draws in the left panels are generated from the DSGE model, whereas the right panels are based on the DSGE-VAR($\hat{\lambda}$). The plots can be compared to Figure 8. The model check discussed in Section 5.1 compares the actual values of sample moments to the posterior predictive distribution of the sample moments. In Figure 13, on the other hand, we compare posterior distributions of sample moments generated from two different model. Since the DSGE-VAR($\hat{\lambda}$) has much higher posterior probability than the DSGE model itself it seems natural to view the corresponding posterior distributions as a benchmark.

How can we formally compare the two posterior distributions? DeJong, Ingram, and Whiteman (1996) propose to use a confidence set overlap criterion, one could compute the Kullback-Leibler divergence between the two posterior distributions, or as in Schorfheide (2000) formulate the following decision problem: Let ζ be a vector of the population characteristics of interest, $\hat{\zeta}$ a point predictor of these population characteristics, and $L(\zeta, \hat{\zeta})$ a loss function. We can generate a point prediction from the DSGE model $\mathcal{M}_1$ by solving

$$\hat{\zeta}_{(1)} = \text{argmin}_{\zeta} \int L(\tilde{\zeta}, \zeta) p(\tilde{\zeta}|Y, \mathcal{M}_1) d\tilde{\zeta}$$

and evaluate the DSGE model based on the posterior risk

$$\int L(\zeta, \hat{\zeta}_{(1)}) p(\zeta|Y, \mathcal{M}_0) d\zeta. \tag{63}$$

The loss-function-based DSGE model evaluation framework in Schorfheide (2000) is more general in the sense that the posterior risk (63) is calculated under the overall posterior distribution of ζ that is obtained by averaging predictions from the reference model $\mathcal{M}_0$ and the DSGE model(s). Moreover, Schorfheide (2000) uses the risk measure to compare multiple, competing DSGE models.

Rather than computing a formal measure of discrepancy we will visually inspect the posteriors. In general, the DSGE-VAR draws are more dispersed than the DSGE model draws reflecting the additional degree of uncertainty introduced by allowing deviations from

the DSGE model restrictions. Most notably, the DSGE-VAR posterior assigns a much higher probability to correlations between inflation and lagged output growth than the DSGE model. This is consistent with the model check reported in Section 5.1 for which the sample correlation of inflation and lagged output growth in the far right tail of the posterior predictive distribution generated by the DSGE model.

6 Nonlinear Methods

For a nonlinear/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 out the latent state vector. The evaluation of the likelihood function can be implemented with a particle filter, also known as sequential Monte Carlo filter. Gordon, Salmond, and Smith (1993) and Kitagawa (1996) are early contributors to this literature. Arulampalam, Maskell, Gordon, and Clapp (2002) provide an excellent survey. In economics, particle filter have been applied to analyze the stochastic volatility models by Pitt and Shephard (1999) and Kim, Shephard and Chib (1998). Recently Fernández-Villaverde and Rubio-Ramírez (2004b) 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', $t = 1, \ldots, \tau$.

Particle Filter

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

2. *Prediction:* Draw one-step ahead forecasted particles $\left\{\tilde{s}_t^i\right\}_{i=1}^N$ from $p\left(s_t|Y^{t-1}, \theta\right)$. Note that

$$p\left(s_t|Y^{t-1}, \theta\right) = \int p\left(s_t|s_{t-1}, \theta\right) p\left(s_{t-1}|Y^{t-1}, \theta\right) ds_{t-1} \approx \frac{1}{N}\sum_{i=1}^{N} p\left(s_t|s_{t-1}^i, \theta\right)$$

   Hence, one can draw draw N particles from $p\left(s_t|Y^{t-1}, \theta\right)$ by generating one particle from $p\left(s_t|s_{t-1}^i, \theta\right)$ for each i.

3. *Filtering:* The goal is to approximate

$$p\left(s_t|Y^t, \theta\right) = \frac{p\left(y_t|s_t, \theta\right) p\left(s_t|Y^{t-1}, \theta\right)}{p\left(y_t|Y^{t-1}, \theta\right)}, \tag{64}$$

which amounts to updating the probability weights assigned to the particles $\left\{\tilde{s}_t^i\right\}_{i=1}^N$. We begin by computing the unnormalized importance weights $\tilde{\pi}_t^i = p\left(y_t|\tilde{s}_t^i,\theta\right)$. The numerator in (64) can be approximated by

$$p\left(y_t|Y^{t-1},\theta\right) = \int p\left(y_t|s_t,\theta\right)\, p\left(s_t|Y^{t-1},\theta\right)\, ds_t \;\approx\; \frac{1}{N}\sum_{i=1}^{N}\tilde{\pi}_t^i. \tag{65}$$

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 $p\left(s_t|\, y^t,\theta\right)$.

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 $p(s_t|\, y^t,\theta)$ given by $\left\{\tilde{s}_t^i,\pi_t^i\right\}_{i=1}^N$ so that

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

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

5. *Likelihood Evaluation:* According to (65) the log likelihood function can be approximated as follows

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

To compare results from first and second-order accurate DSGE model solutions we simulated an artificial data of 80 observations from the quadratic approximation of model $\mathcal{M}_1$ with parameter values reported as DGP 3 in Table 1. In the remainder of this section we will refer to these parameters as 'true' values as opposed to 'pseudo-true' values to be defined later. These true parameter values are chosen to match the first and second moments of the simulated series to the sample moments of Data Set 1-$\mathcal{M}_1$ generated from the linear approximation of $\mathcal{M}_1$. The quadratic approximation introduces non-zero intercepts $\Phi_j^{(0)}$ into the state transition equations (36). Since for our model these intercepts are negative, the steady state inflation and real interest rates have to be adjusted upwards.

[12]The Sequential Importance Sampler (SIS) is a Monte Carlo method which utilizes this aspect, but it is well documented to suffer from a degeneracy problem, where after a few iterations, all but one particle will have negligible weight.

Two additional parameters enter the quadratic approximation: ν and $1/g$. ν is set to 0.15 which is close to the Laforte's (2003) calibration value, 1/7, and $1/g$ is chosen as 0.85, which implies that steady state government consumption amounts to 15% of output. For computational convenience, a measurement error of standard deviation 0.1 is introduced to each measurement equation.[13]

6.1 Configuration of the Particle Filter

There are several issues on the practical implementation of the particle filter. First, we need a scheme to draw from the initial state distribution, $p(s_0|\theta)$. In the linear case, it is straightforward to calculate the unconditional distribution of s_t associated with the vector autoregressive representation (32). For the second-order approximation we rewrite the state transition equation (36) as follows. Decompose $s_t = [x_t', z_t']'$, where z_t (not to be confused with the technology growth process in Section 2) is composed of the exogenous processes $\epsilon_{R,t}$, $\hat{g}_t$, and $\hat{z}_t$. The law of motion for the endogenous state variables $x_{j,t}$ can be expressed as

$$x_{j,t} = F_j^{(0)} + F_j^{(1)} w_t + w_t' F^{(2)} w_t \tag{66}$$

where $w_t = [x_{t-1}', z_t']'$. We generate x_0 by drawing z_0 from its unconditional distribution (the law of motion for z_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 $p(s_t|Y^t)$. 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. Fernández-Villaverde and Rubio-Ramírez (2004b) show by simulation that the gain from more particles is very marginal after 20,000 particles in their specification, and it motivates our choice of 40,000 particles. If the measurement errors in the conditional distribution $p(y_t|s_t)$ 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.

[13] Without the measurement error two complications arise: since $p(y_t|s_t)$ degenerates to a point-mass and the distribution of s_t is that of a multivariate non-central χ^2 the evaluation of $p(y_t|Y^{t-1})$ becomes difficult. Moreover, the computation of $p(z_t|Y^t)$ requires the solution of a system of quadratic equations.

Even though the procedure involves extensive random sampling, the particle filter can be readily implementable on a good desktop computer. We implement most of the procedure in MATLAB 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.

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. Since we had to adjust the parameter values to generate data from the second-order approximation of $\mathcal{M}_1$ that have similar (unconditional) first and second moments as the data drawn from the linear approximation it is not surprising that the corresponding likelihood functions peak in different regions of the DSGE model parameter space. It is natural to evaluate the quadratic likelihood function in the neighborhood of the true parameter values given in Table 1. However, we evaluate the linear likelihood around pseudo-true parameter values, which are obtained by finding the mode of the linear likelihood using 3,000 observations from DGP 3. 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 (31) because ν and φ cannot be identified with the log-linearized DSGE model. To match κ between DGP 1 and DGP 3 we set $\varphi = 74.4023$.

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 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 $\pi^{(A)}$. This finding has first-order welfare implications in the context of this of the model. As noted in An (2005), an evaluation of monetary policy should be based on nonlinear estimation results rather than on linear estimates plugged into a second-order approximated DSGE model as it has been done in the literature. Another striking feature is that the quadratic log-likelihood function is more concave than the linear one, which implies that the nonlinear approach is able to extract more information on the structural parameters from the data. For instance, it appears that the monetary

policy parameters, ψ_1 and ψ_2 can be much 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. $1/g$ is dropped in the linear version of the model and hence the log-likelihood profile along this dimension is flat. In the quadratic approximation, however, we can see that the log-likelihood is concave in $1/g$. Moreover, the linear likelihood is flat for values of ν and φ that imply a constant κ. This is illustrated in Figure 15 which depicts 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 shows that the iso-κ lines (dashed) intersect with the contours of the quadratic likelihood function, which suggests that ν and φ are separately identifiable.

6.3 The Posterior

We now compare the posterior distributions obtained from the linear and nonlinear analysis. In both cases we use the same prior distribution which is reported in Table 1. Unlike in the analysis of the likelihood function we now substitute the adjustment cost parameter φ by the Phillips-curve parameter κ in the quadratic approximation of the DSGE model. A fairly uninformative beta prior with mean 0.5 and standard deviation 0.20 is used for ν since little empirical evidence on this parameter has been provided in the literature. 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 0.1 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 nonlinear analysis. We use scaling factors 0.5 (linear) and 0.025 (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.

Figure 16 and Figure 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. As expected, the draws appear to be more concentrated as we move from

prior to linear posterior, and further to quadratic posterior. Since the linear likelihood is flatter than the quadratic likelihood, the linear posterior is more sensitive to the specification of the prior. For instance, the linear posteriors of τ, κ, σ_R and ρ_g are shifted towards the prior means, away from the pseudo-true values. For quadratic posterior, on the other hand, the shift away from the true parameter values is less pronounced since the likelihood is much more informative and the prior does not have a great effect on the posterior.

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 nonlinear 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 nonlinear 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

Adolfson, Malin, Stefan Laséen, Jesper Lindé, and Mattias Villani (2004): "Bayesian Estimation of an Open Economy DSGE Model with Incomplete Pass-Through," *Manuscript*, Sveriges Riksbank.

Altug, Sumru (1989): "Time-to-Build and Aggregate Fluctuations: Some New Evidence," *International Economic Review*, **30**, 889-920.

An, Sungbae (2005): "Evaluating Welfare Implications of Monetary Policy Rules Based on an Estimation of Second-Order Approximated Model," *Manuscript*, University of Pennsylvania.

Anderson, Gary (2000): "A Reliable and Computationally Efficient Algorithm for Imposing the Saddle Point Property in Dynamic Models," *Manuscript*, Federal Reserve Board of Governors, Division of Monetary Affairs.

Arulampalam, Sanjeev, Simon Maskell, Neil Gordon, and Tim Clapp (2002): "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking," *IEEE Transactions on Signal Processing*, **50**, 174-188.

Aruoba, Borağan, Jesús Fernandéz-Villaverde, and Juan Rubio-Ramırez (2004): "Comparing Solution Methods for Dynamic Equilibrium Economies," *Manuscript*, University of Pennsylvania.

Beyer, Andreas and Roger Farmer (2004): "On the Indeterminacy of New Keynesian Economics," *ECB Working Paper*, **323**.

Bierens, Herman (2005): "Econometric Analysis of Linearized Singular Dynamic Stochastic General Equilibrium Models," *Journal of Econometrics*, forthcoming.

Bils, Mark and Peter Klenow (2004): "Some Evidence on the Importance of Sticky Prices," *Journal of Political Economy*, **112**, 947-985.

Binder, Michael and Hashem Pesaran (1997): "Multivariate Linear Rational Expectations Models: Characterization of the Nature of the Solutions and Their Fully Recursive Computation," *Econometric Theory*, **13**, 877-888.

Blanchard, Olivier and Charles Kahn (1980): "The Solution of Linear Difference Models under Rational Expectations," *Econometrica*, **48**, 1305-1311.

Box, George (1980): "Sampling and Bayes' Inference in Scientific Modelling and Robustness," *Journal of the Royal Statistical Society A*, **143**, 383-430.

Canova, Fabio (1994): "Statistical Inference in Calibrated Models," *Journal of Applied Econometrics*, **9**, S123-S144.

Canova, Fabio (2004): "Monetary Policy and the Evolution of US Economy: 1948-2002," *Manuscript*, IGIER and Bocconi University.

Canova, Fabio (2005): *"Methods for Applied Macroeconomic Research,"* Princeton University Press, forthcoming.

Canova, Fabio and Luca Sala (2005): "Back to Square One: Identification Issues in DSGE Models," *Manuscript*, IGIER and Bocconi University.

Chang, Yongsung, Joao Gomes, and Frank Schorfheide (2002): "Learning-by-Doing as Propagation Mechanism," *American Economic Review*, **92**, 1498-1520.

Chang, Yongsung and Frank Schorfheide (2003): "Labor-Supply Shifts and Economic Fluctuations," *Journal of Monetary Economics*, **50**, 1751-1768.

Chib, Siddhartha and Edward Greenberg (1995): "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, **49**, 327-335.

Chib, Siddhartha and Ivan Jeliazkov (2001): "Marginal Likelihood from the Metropolis-Hastings Output," *Journal of the American Statistical Association*, **96**, 270-281.

Christiano, Lawrence (2002): "Solving Dynamic Equilibrium Models by a Method of Undetermined Coefficients," *Computational Economics*, **20** (1-2).

Christiano, Lawrence and Martin Eichenbaum (1992): "Current Real-Business Cycle Theories and Aggregate Labor-Market Fluctuations," *American Economic Review*, **82**, 430-450.

Christiano, Lawrence , Martin Eichenbaum, and Charles Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, **113**, 1-45.

Collard, Fabrice and Michel Juillard (2001): "Perturbation Methods for Rational Expectations Models," *Manuscript*, CEPREMAP.

Crowder, Martin (1988): "Asymptotic Expansions of Posterior Expectations, Distributions, and Densities for Stochastic Processes," *Annals of the Institute for Statistical Mathematics*, **40**, 297-309.

DeJong, David and Beth Ingram (2001): "The Cyclical Behavior of Skill Acquisition," *Review of Economic Dynamics*, **4**, 536-561.

DeJong, David, Beth Ingram, and Charles Whiteman (1996): "A Bayesian Approach to Calibration," *Journal of Business Economics and Statistics*, **14**, 1-9.

DeJong, David, Beth Ingram, and Charles Whiteman (2000): "A Bayesian Approach to Dynamic Macroeconomics," *Journal of Econometrics*, **98**, 201-223.

Del Negro, Marco (2003): "Fear of Floating: A Structural Estimation of Monetary Policy in Mexico," *Manuscript*, Federal Reserve Bank of Atlanta.

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

Del Negro, Marco, Frank Schorfheide, Frank Smets, and Raf Wouters (2004): "On the Fit and Forecasting Performance of New Keynesian Models," *Manuscript*, University of Pennsylvania.

Den Haan, Wouter and Albert Marcet (1994): "Accuracy in Simulations," *Review of Economic Studies*, **61**, 3-18.

De Walque, Grégory and Raf Wouters (2004): "An Open Economy DSGE Model Linking the Euro Area and the U.S. Economy," *Manuscript*, National Bank of Belgium.

Diebold, Francis, Lee Ohanian, and Jeremy Berkowitz (1998): "Dynamic Equilibrium Economies: A Framework for Comparing Models and Data," *Review of Economic Studies*, **65**, 433-452.

Dridi, Ramdan, Alain Guay, and Eric Renault (2004): "Calibration of DSGE Models: An Sequential Partial Indirect Inference Approach," *Manuscript*, University of North Carolina.

DYNARE: software available at *http://www.cepremap.cnrs.fr/dynare/*.

Fernández-Villaverde, Jesús and Juan Rubio-Ramírez (2004a): "Comparing Dynamic Equilibrium Models to Data: a Bayesian Approach," *Journal of Econometrics*, **123**, 153-187.

Fernández-Villaverde, Jesús and Juan Rubio-Ramírez (2004b): "Estimating Macroeconomic Models: A Likelihood Approach," *Manuscript*, University of Pennsylvania.

Galí, Jordi and Pau Rabanal (2004): "Technology Shocks and Aggregate Fluctuations: How Well Does the RBC Model Fit of Postwar U.S. Data" in Mark Gertle and Kenneth Rogoff (eds.) *NBER Macroeconomics Annual 2004*, MIT Press, Cambridge, MA.

GAUSS: software available from : *http://www.aptech.com/*.

Gelfand, Alan and Dipak Dey (1994): "Bayesian Model Choice: Asymptotics and Exact Calculations," *Journal of the Royal Statistical Society B*, **56**, 501-514.

Gelman, Andrew, John Carlin, Hal Stern, and Donald Rubin (1995): *"Bayesian Data Analysis,"* Chapman & Hall, New York.

Geweke, John (1989): "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, **57**, 1317-1340.

Geweke, John (1999a): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication," *Econometric Reviews*, **18**, 1-126.

Geweke, John (1999b): "Computational Experiments and Reality," *Manuscript*, University of Iowa.

Good, Irving (1952): "Rational Decisions," *Journal of the Royal Statistical Society B*, **14**, 107-114.

Gordon, Neil, David Salmond, and Adrian Smith (1993): "Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation," *IEEE Proceedings-F*, **140**, 107-113.

Hammersley, John and David Handscomb (1964): *"Monte Carlo Methods,"* John Wiley, New York.

Hansen, Lars and James Heckman (1996): "The Empirical Foundations of Calibration," *Journal of Economic Perspectives*, **10**, 87-104.

Hastings, Keith (1970): "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, **57**, 97-109.

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

Ireland, Peter (2004): "A Method for Taking Models to the Data," *Journal of Economic Dynamics and Control*, **28**, 1205-1226.

Jin, He-Hui and Kenneth Judd (2002): "Perturbation Methods for General Dynamic Stochastic Models," *Manuscript*, Hoover Institute.

Judd, Kenneth (1998): "*Numerical Methods in Economics*," MIT Press, Cambridge, MA.

Justiniano, Alejandro and Bruce Preston (2004): "Small Open Economy DSGE Models – Specification, Estimation, and Model Fit," *Manuscript*, International Monetary Fund and Department of Economics, Columbia University.

Kass, Robert and Adrian Raftery (1995): "Bayes Factors," *Journal of the American Statistical Association*, **90**,773-795.

Kim, Jae-Young (1998): "Large Sample Properties of Posterior Densities, Bayesian Information Criterion, and the Likelihood Principle in Nonstationary Time Series Models," *Econometrica*, **66**, 359-380.

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

Kim, Jinill, and Sunghyun Kim (2003): "Spurious Welfare Reversals in International Business Cycle Models," *Journal of International Economics*, **60**, 471-500.

Kim, Jinill, Sunghyun Kim, Ernst Schaumburg, and Christopher Sims (2005) "Calculating and Using Second Order Accurate Solutions of Discrete Time Dynamic Equilibrium Models," *Manuscript*, Princeton University.

Kim, Kunhong and Adrian Pagan (1995): "Econometric Analysis of Calibrated Macroeconomic Models" in Hashem Pesaran and Mike Wickens (eds.) *"Handbook of Applied Econometrics: Macroeconomics,"* Basil Blackwell, Oxford.

Kim, Sangjoon, Neil Shephard, and Siddhartha Chib (1998): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH models," *Review of Economic Studies*, **65**, 361-93.

King, Robert and Mark Watson (1998): "The Solution of Singular Linear Difference Systems under Rational Expectations," *International Economic Review*, **39**, 1015-1026.

Kitagawa Genshiro (1996): "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models," *Journal of Computational & Graphical Statistics*, **5**, 1-25.

Klein, Paul (2000): "Using the Generalized Schur Form to Solve a Multivariate Linear Rational Expectations Model," *Journal of Economic Dynamics and Control*, **24**, 1405-1423.

Kydland, Finn and Edward Prescott (1982): "Time to Build and Aggregate Fluctuations," *Econometrica*, **50**, 1345-70.

Kydland, Finn and Edward Prescott (1996): "The Computational Experiment: An Econometric Tool," *Journal of Economic Perspecitives*, **10**, 69-85.

Laforte, Jean-Philippe (2004): "Comparing Monetary Policy Rules in an Estimated General Equilibrium Model of the U.S. Economy," *Manuscript*, Federal Reserve Board of Governors.

Landon-Lane, John (1998): "Bayesian Comparison of Dynamic Macroeconomic Models," *Ph.D. Dissertation*, University of Minnesota.

Leeper, Eric and Christopher Sims (1994): "Toward a Modern Macroeconomic Model Usable for Policy Analysis" in Stanley Fischer and Julio Rotemberg (eds.) *NBER Macroeconomics Annual 1994*, MIT Press, Cambridge, MA.

Levin, Andrew, Alexei Onatski, John Williams, and Noah Williams (2005): "Monetary Policy Under Uncertainty in Micro-Founded Macroeconometric Models" in Mark Gertler and Kenneth Rogoff (eds.) *NBER Macroeconomics Annual 2005*, MIT Press, Cambridge, MA.

Lubik, Thomas and Frank Schorfheide (2003): "Do Central Banks Respond to Exchange Rate Fluctuation – A Structural Investigation," *Manuscript*, University of Pennsylvania.

Lubik, Thomas and Frank Schorfheide (2004): "Testing for Indeterminacy: An Application to U.S. Monetary Policy," *American Economic Review*, **94**, 190-217.

Lubik, Thomas and Frank Schorfheide (2005): "A Bayesian Look at New Open Economy Macroeconomics" in Mark Gertler and Kenneth Rogoff (eds.) *NBER Macroeconomics Annual 2005*, MIT Press, Cambridge, MA.

MATLAB: software available from *http://www.mathworks.com/*.

McGrattan, Ellen (1994): "The Macroeconomic Effects of Distortionary Taxation," *Journal of Monetary Economics*, **33**, 573-601.

Metropolis, Nicholas, Arianna Rosenbluth, Marshall Rosenbluth, Augusta Teller, and Edward Teller (1953): "Equations of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, **21**, 1087-1092.

Onatski, Alexei and Noah Williams (2004): "Empirical and Policy Performance of a Forward-looking Monetary Model," *Manuscript*, Columbia University.

Otrok, Christopher (2001): "On Measuring the Welfare Cost of Business Cycles," *Journal of Monetary Economics*, **47**, 61-92.

Phillips, Peter (1996): "Econometric Model Determination," *Econometrica*, **64**, 763-812.

Pitt, Michael and Neil Shephard (1999): "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, **94**, 590-599.

Poirier, Dale (1998): "Revising Beliefs in Nonidentified Models," *Econometric Theory*, **14**, 183-209.

Rabanal, Pau and Juan Rubio-Ramírez (2003): "Comparing New Keynesian Models in the Euro Area: A Bayesian Approach," *Federal Reserve Bank of Atlanta Working Paper*, **2003-30**.

Rabanal, Pau and Juan Rubio-Ramírez (2005): "Comparing New Keynesian Models of the Business Cycle: A Bayesian Approach," *Journal of Monetary Economics*, forthcoming.

Rabanal, Pau and Vicente Tuesta (2005): "Euro-Dollar Real Exchange Rate Dynamics in an Estimated Two-Country Model: What is Important and What is Not," *Manuscript*, International Monetary Fund.

Robert, Christian and George Casella (1999): *"Monte Carlo Statistical Methods,"* Springer Verlag, New York.

Rotemberg, Julio and Michael Woodford (1997): "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy" in Ben Bernanke and Julio Rotemberg (eds.) *NBER Macroeconomics Annual 1997*, MIT Press, Cambridge, MA.

Sargent, Thomas (1989): "Two Models of Measurements and the Investment Accelerator," *Journal of Political Economy*, **97**, 251-287.

Schmitt-Grohé, Stephanie and Martín Uribe (2004): "Solving Dynamic General Equilibrium Models Using a Second-Order Approximation to the Policy Function," *Journal of Economic Dynamics and Control*, **28**, 755-775.

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

Schorfheide, Frank (2005): "Learning and Monetary Policy Shifts," *Review of Economic Dynamics*, **8**, 392-419.

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

Sims, Christopher (2002): "Solving Linear Rational Expectations Models," *Computational Economics*, **20**, 1-20.

Sims, Christopher A. (2003): "Probability Models for Monetary Policy Decisions," *Manuscript*, Princeton University.

Smets, Frank and Raf Wouters (2003): "An Estimated Stochastic Dynamic General Equilibrium Model for the Euro Area," *Journal of the European Economic Association*, **1**, 1123-1175.

Smets, Frank and Raf Wouters (2005): "Comparing Shocks and Frictions in U.S. and Euro Area Business Cycles: A Bayesian DSGE Approach," *Journal of Applied Econometrics*, **20**, 161-183.

Smith, Anthony (1993): "Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, **8**, S63-S84.

Swanson, Eric, Gary Anderson, and Andrew Levin (2005): "Higher-Order Perturbation Solutions to Dynamic Discrete-Time Rational Expectations Models," *Manuscript*, Federal Reserve Board of Governors.

Taylor, John and Harald Uhlig (1990): "Solving Non-linear Stochastic Growth Models: A comparison of Alternative Solution Methods," *Journal of Business Economics and Statistics*, **8**, 1-34.

Uhlig, Harald (1999): "A Toolkit for Analyzing Nonlinear Dynamic Stochastic Models Easily" in Ramon Marimóm and Adrian Scott (eds.) *Computational Methods for the Study of Dynamic Economies*, Oxford University Press.

Walker, A.M. (1969): "On the Asymptotic Behaviour of Posterior Distributions," *Journal of the Royal Statistical Society B*, **31**, 80-88.

Watson, Mark (1993): "Measures of Fit for Calibrated Models," *Journal of Political Economy*, **101**, 1011-1041.

White, Halbert (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, **50**, 1-25.

Woodford, Michael (2003): "*Interest and Prices*," Princeton University Press, Princeton.

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

Table 1: Prior Distribution and DGP

Name	Domain	Density	Para (1)	Para (2)	DGP 1	DGP 3
τ	$\mathbb{R}^+$	Gamma	2.00	0.50	2.00	2.00
κ	$\mathbb{R}^+$	Gamma	0.20	0.10	0.15	0.15
ψ_1	$\mathbb{R}^+$	Gamma	1.50	0.25	1.50	1.50
ψ_2	$\mathbb{R}^+$	Gamma	0.50	0.25	1.00	1.00
ρ_R	$[0,1)$	Beta	0.50	0.20	0.60	0.60
ρ_G	$[0,1)$	Beta	0.80	0.10	0.95	0.95
ρ_Z	$[0,1)$	Beta	0.66	0.15	0.65	0.65
$r^{(A)}$	$\mathbb{R}^+$	Gamma	0.50	0.50	0.40	1.00
$\pi^{(A)}$	$\mathbb{R}^+$	Gamma	7.00	2.00	4.00	6.20
$\gamma^{(Q)}$	$\mathbb{R}$	Normal	0.40	0.20	0.50	0.50
σ_R	$\mathbb{R}^+$	InvGamma	0.40	4.00	0.20	0.20
σ_G	$\mathbb{R}^+$	InvGamma	1.00	4.00	0.80	0.80
σ_Z	$\mathbb{R}^+$	InvGamma	0.50	4.00	0.45	0.30
ν	$[0,1)$	Beta	0.50	0.20		0.15
$1/g$	$[0,1)$	Beta	0.85	0.10		0.85

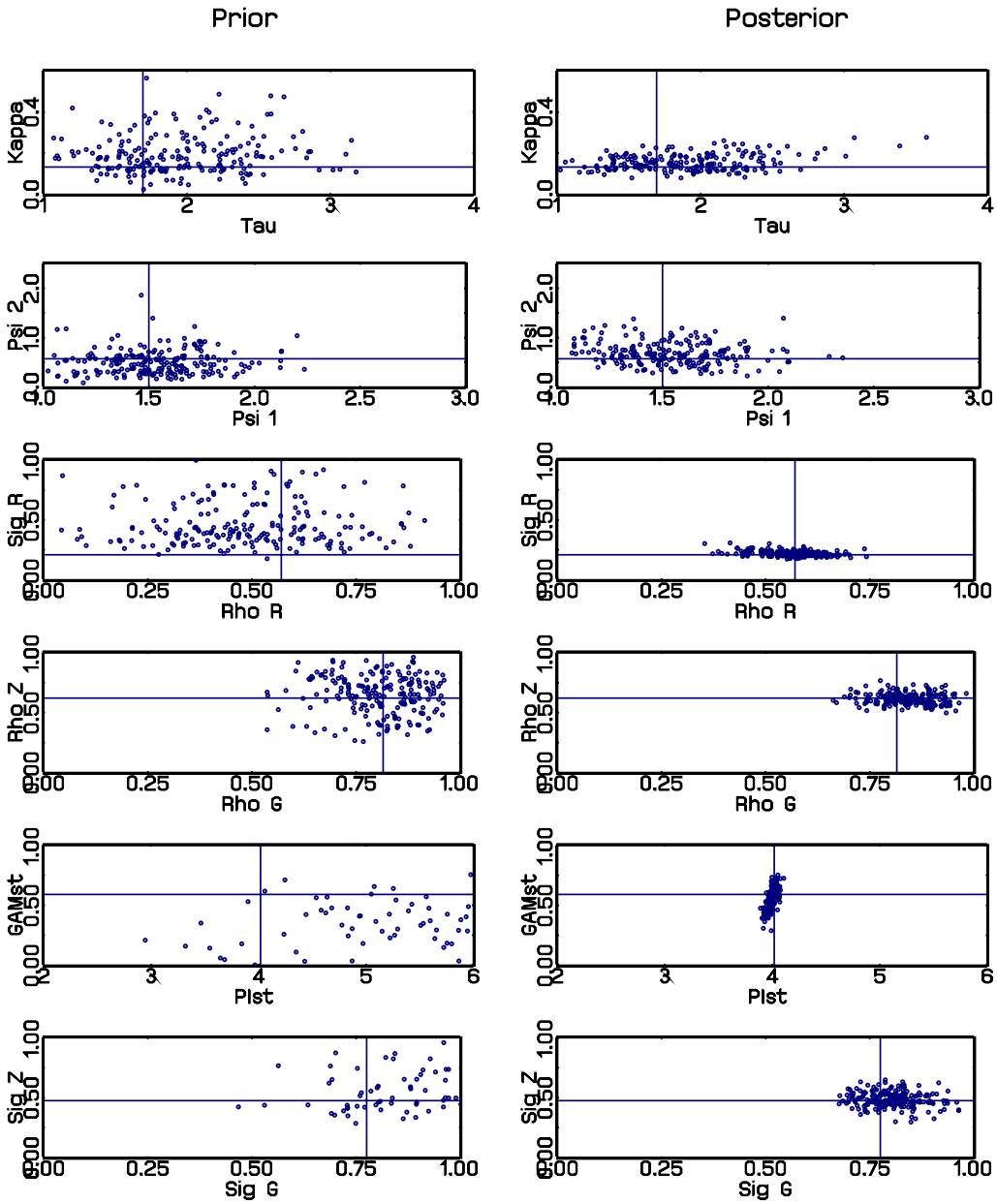
Notes: Para (1) and Para (2) list the means and the standard deviations for Beta, Gamma, and Normal distributions; the upper and lower bound of the support for the Uniform distribution; s and ν for the Inverse Gamma distribution, where $p_{\mathcal{IG}}(\sigma|\nu,s) \propto \sigma^{-\nu-1}e^{-\nu s^2/2\sigma^2}$. The effective prior is truncated at the boundary of the determinacy region. Parameters ν and $1/g$ only affect the second-order accurate solution of the DSGE model.

Table 2: Log Marginal Data Densities

Specification	Data Set 1	Data Set 2
DSGE Model	-196.66	-279.38
DSGE-VAR $\lambda = \infty$	-196.88	-277.49
DSGE-VAR $\lambda = 5.00$	-198.87	-270.46
DSGE-VAR $\lambda = 1.00$	-206.57	-258.25
DSGE-VAR $\lambda = 0.75$	-209.53	-257.53
DSGE-VAR $\lambda = 0.50$	-215.06	-258.73
DSGE-VAR $\lambda = 0.25$	-231.20	-269.66

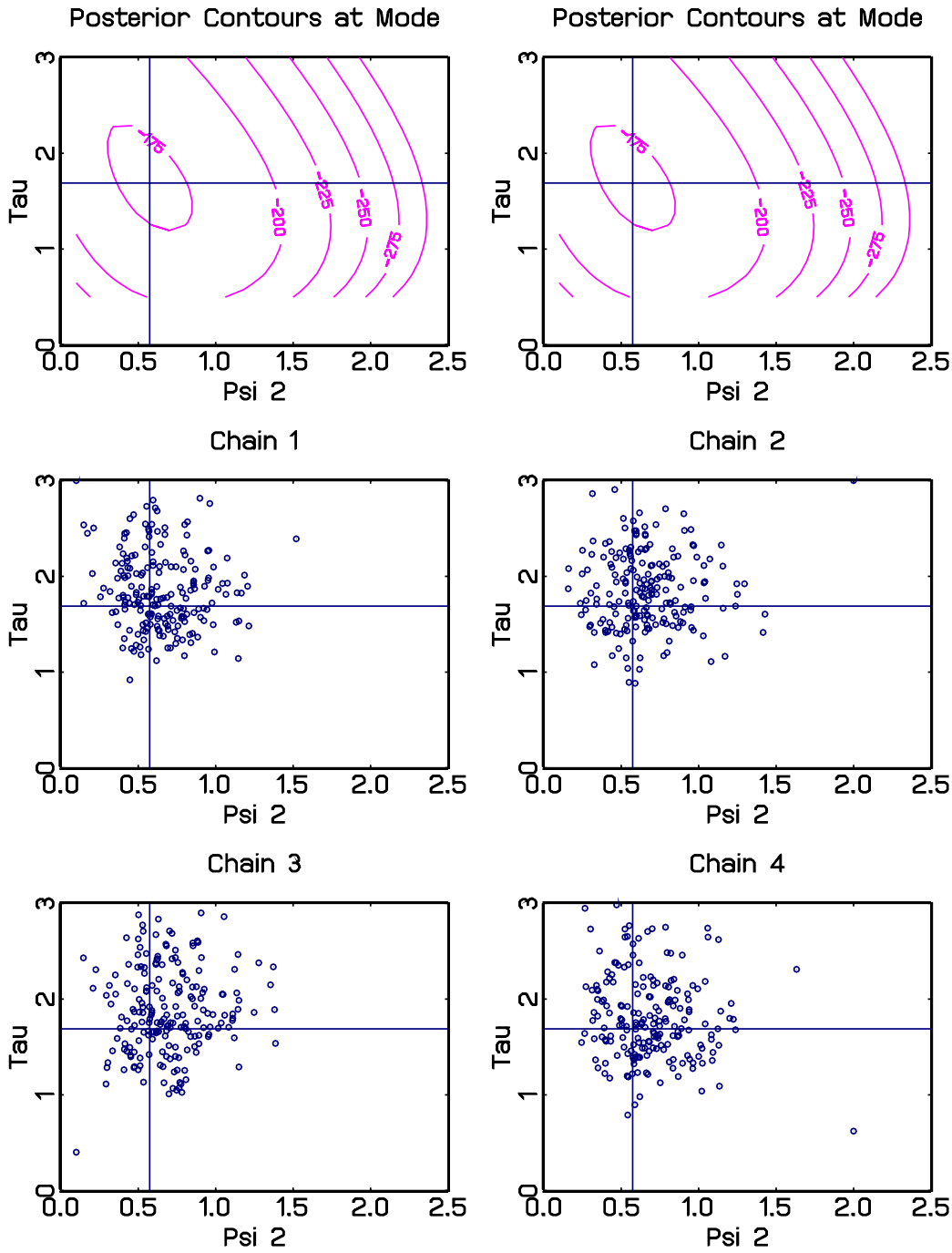
Notes: Output gap rule specification $\mathcal{M}_1$. The log marginal data densities for the DSGE model specifications are computed based on Geweke's (1999) modified harmonic mean estimator.

Figure 1: Prior and Posterior



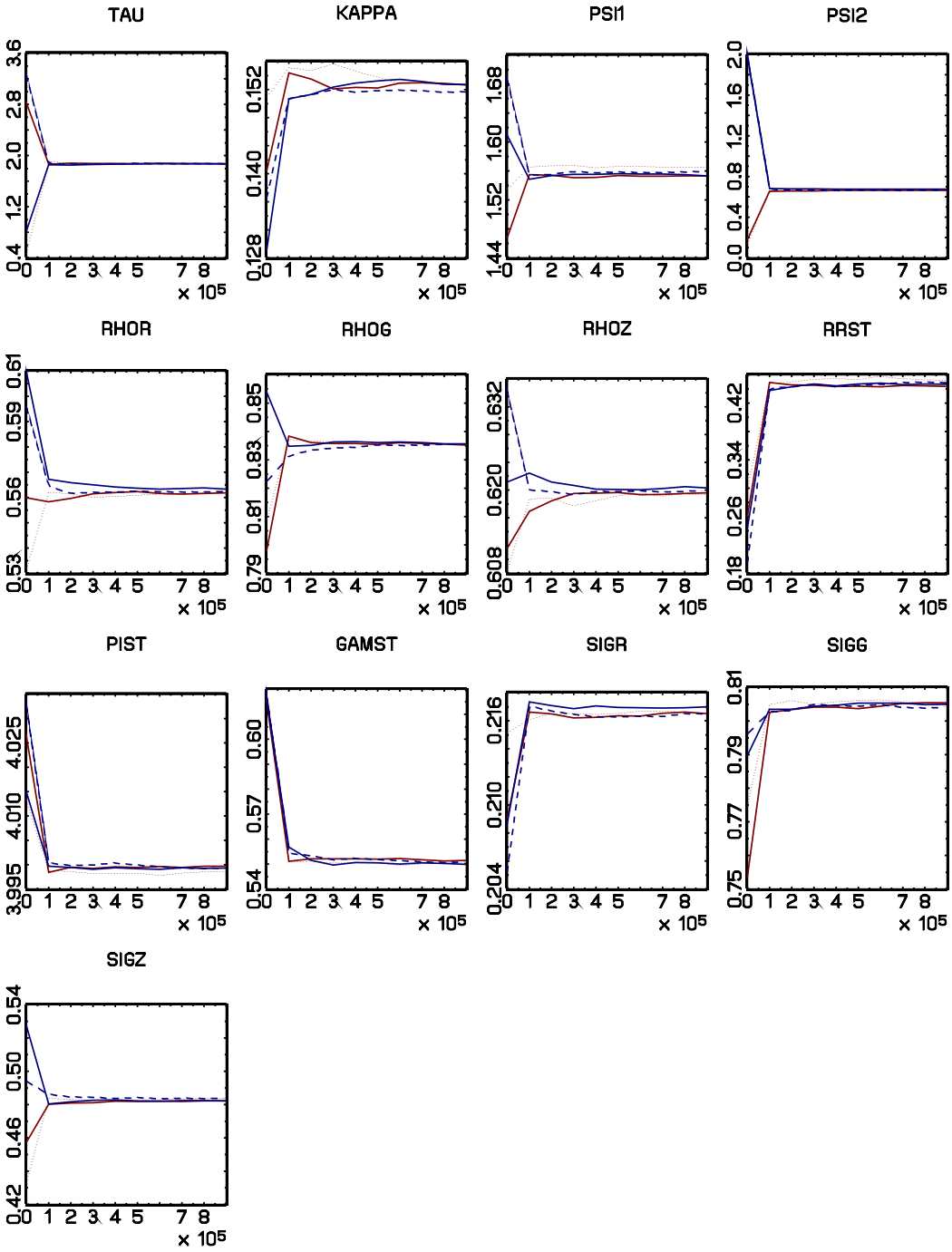
Notes: Output gap rule specification $\mathcal{M}_1$, Data Set 1-$\mathcal{M}_1$. The panels depict 200 draws from prior and posterior distributions. Intersections of solid lines signify posterior mode values.

Figure 2: Draws from Multiple Chains



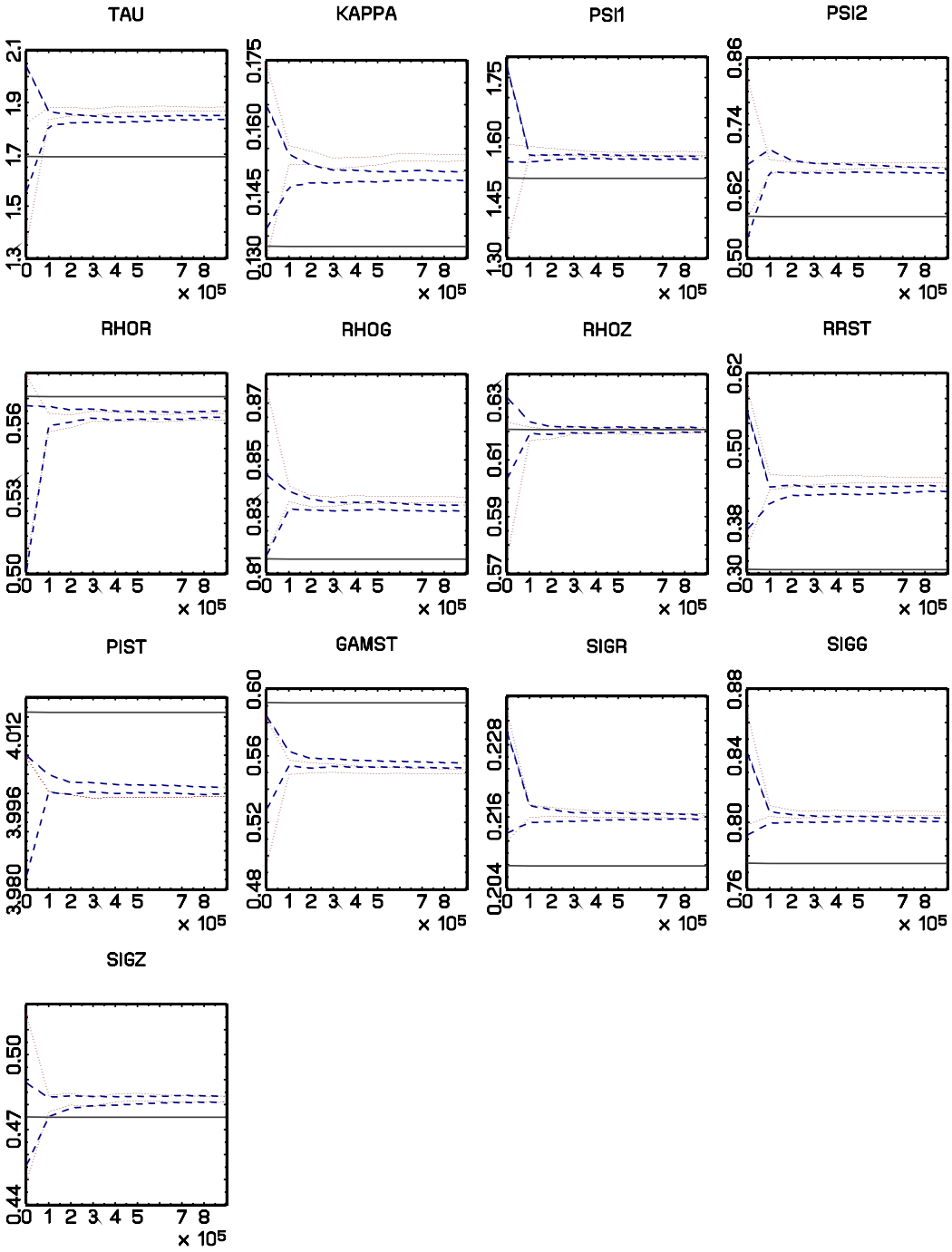
Notes: Output gap rule specification $\mathcal{M}_1$, Data Set 1-$\mathcal{M}_1$. Panels (1,1) and (1,2): contours of posterior density at the mode as function of τ and ψ_2. Panels (2,1) to (3,2): 200 draws from four Markov chains generated by the Metropolis Algorithm. Intersections of solid lines signify posterior mode values.

Figure 3: Recursive Means from Multiple Chains



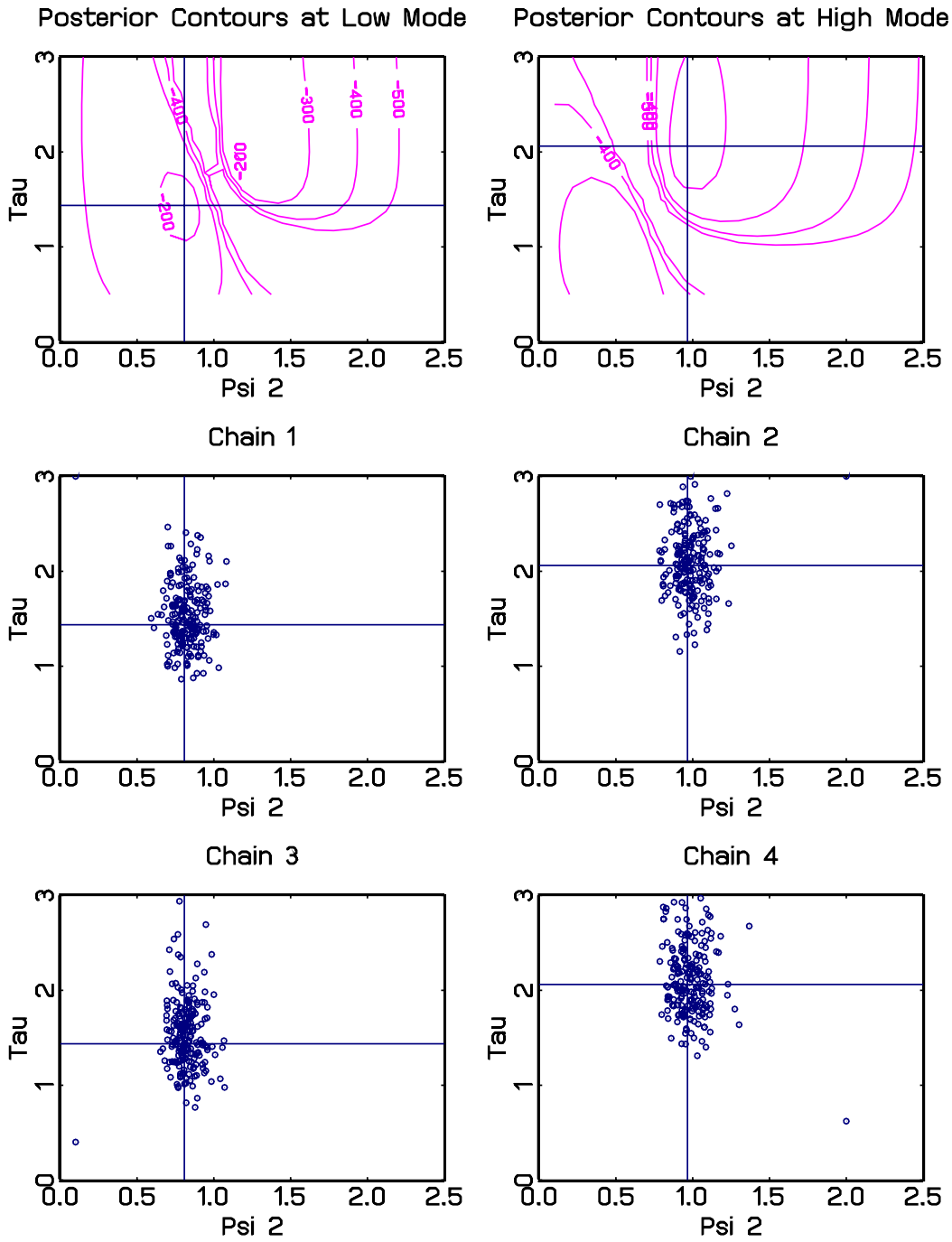
Notes: Output gap rule specification $\mathcal{M}_1$, Data Set 1-$\mathcal{M}_1$. Each line corresponds to recursive means (as a function of the number of draws) calculated from one of the four Markov chains generated by the Metropolis Algorithm.

Figure 4: Metropolis Algorithm versus Importance Sampling



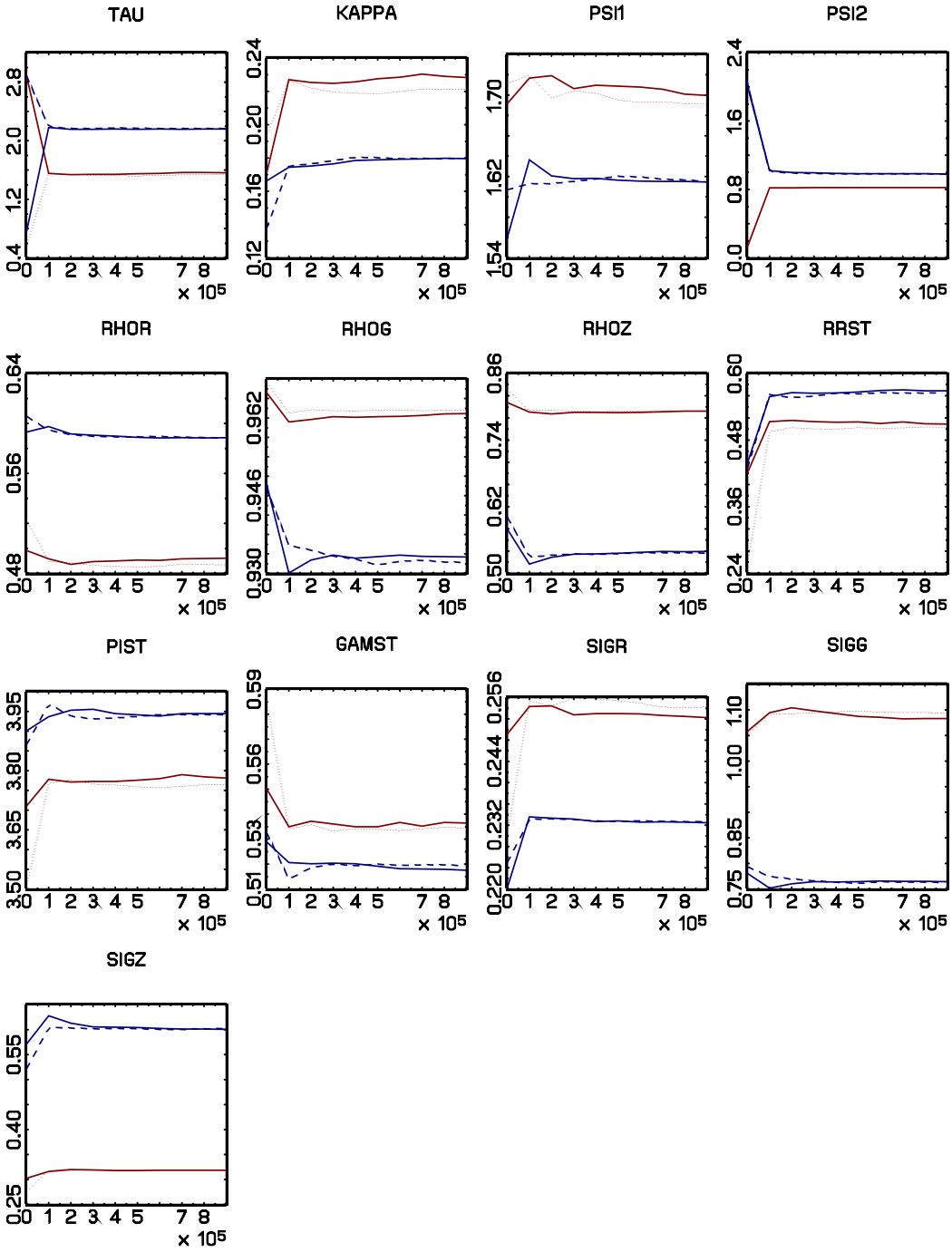
Notes: Output gap rule specification $\mathcal{M}_1$, Data Set 1-$\mathcal{M}_1$. Panels depict posterior modes (solid), recursively computed 95% bands for posterior means based on the Metropolis Algorithm (dotted) and the Importance Sampler (dashed).

Figure 5: Draws from Multiple Chains



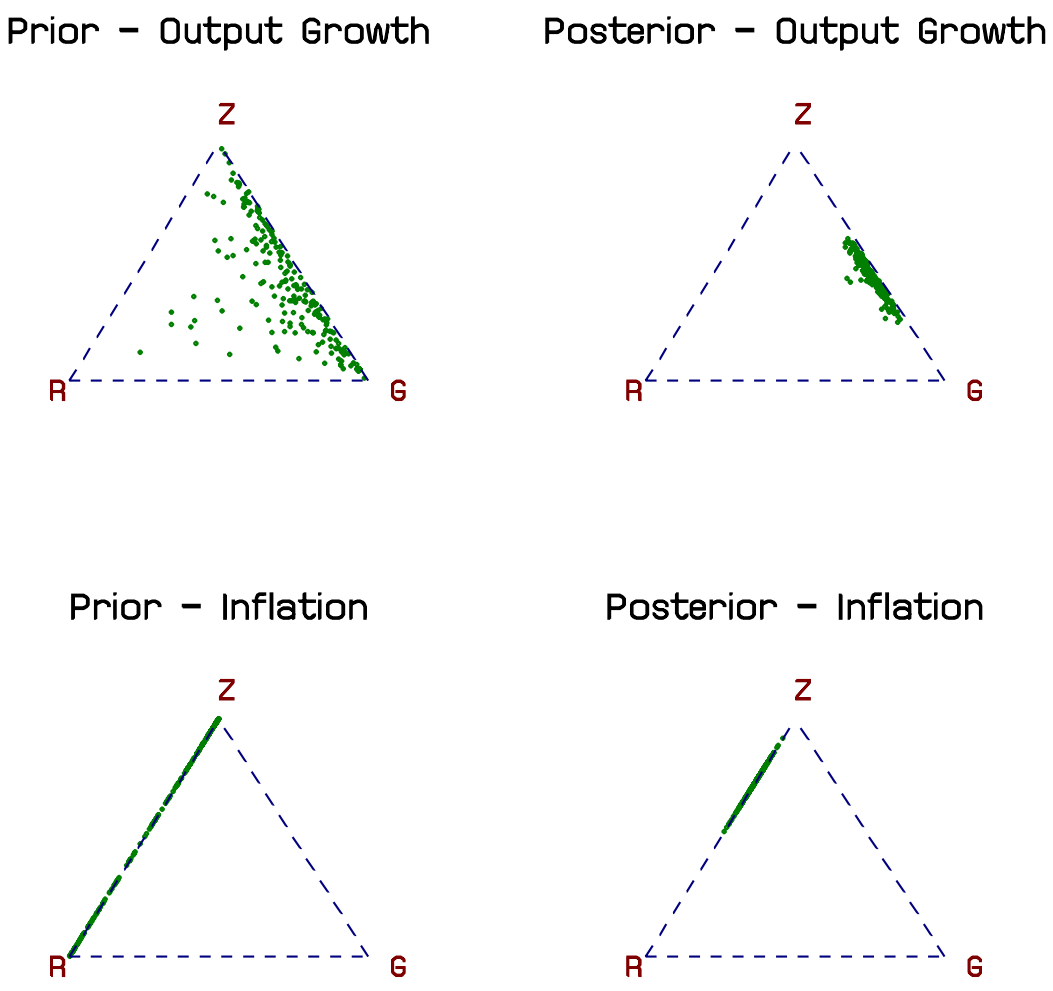
Notes: Output growth rule specification $\mathcal{M}_2$, Data Set 1-$\mathcal{M}_2$. Panels (1,1) and (1,2): contours of posterior density at "low" and "high" mode as function of τ and ψ_2. Panels (2,1) to (3,2): 200 draws from four Markov chains generated by the Metropolis Algorithm. Intersections of solid lines signify "low" (left panels) and "high" (right panels) posterior mode values.

Figure 6: Recursive Means from Multiple Chains



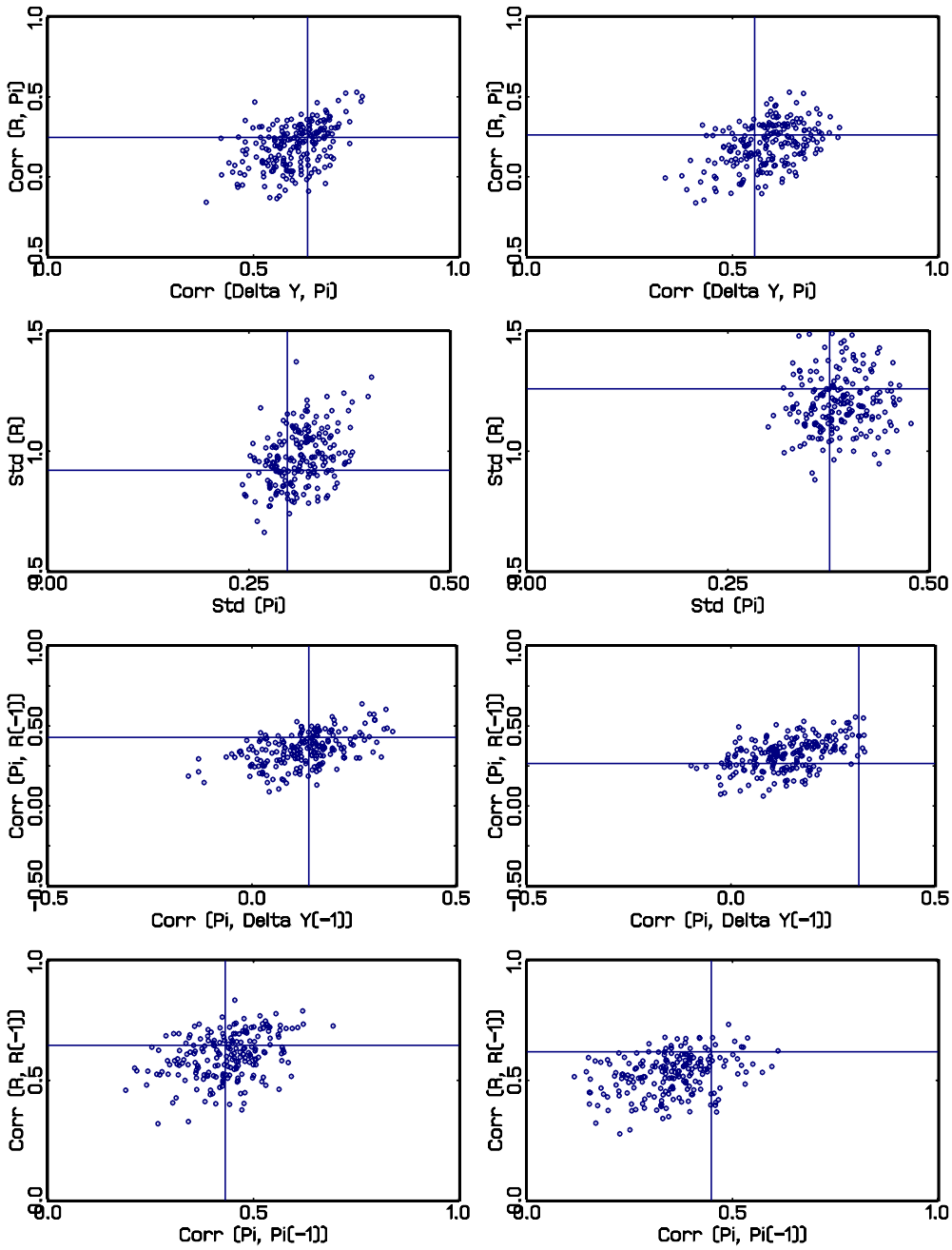
Notes: Output growth rule specification $\mathcal{M}_2$, Data Set 1-$\mathcal{M}_2$. Each line corresponds to recursive means (as a function of the number of draws) calculated from one of the four Markov chains generated by the Metropolis Algorithm.

Figure 7: Prior and Posterior Variance Decompositions



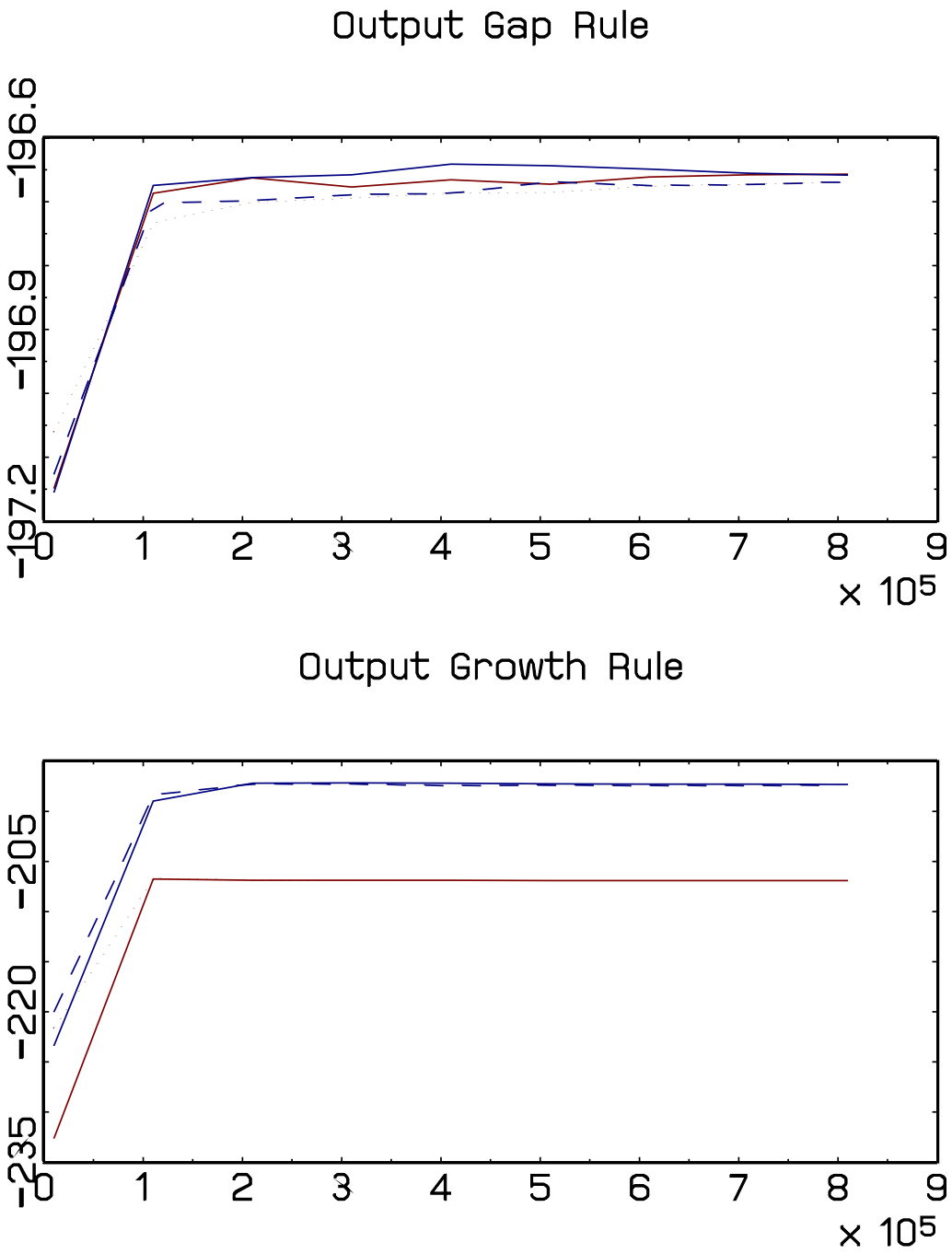
Notes: Output growth rule specification $\mathcal{M}_1$, Data Set 1-$\mathcal{M}_1$. The panels depict 200 draws from prior and posterior distributions arranged on a 3-dimensional simplex. The three corners (Z,G,R) correspond to 100% of the variation being due to the shocks $\epsilon_{z,t}$, $\epsilon_{g,t}$, and $\epsilon_{R,t}$, respectively.

Figure 8: Posterior Predictive Check



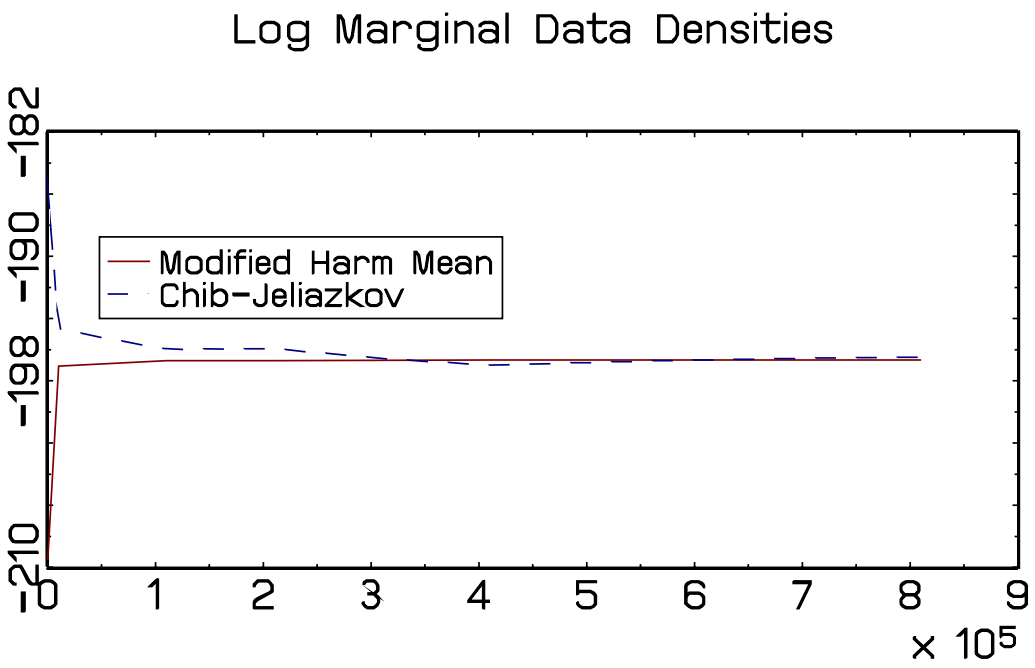
Notes: Output gap rule specification $\mathcal{M}_1$. We plot 200 draws from the posterior predictive distribution of various sample moments. Intersections of solid lines signify the observed sample moments. Left panels: Data Set 1-$\mathcal{M}_1$, no misspecification. Right panels: Data Set 2, misspecification.

Figure 9: Log Marginal Data Densities from Multiple Chains



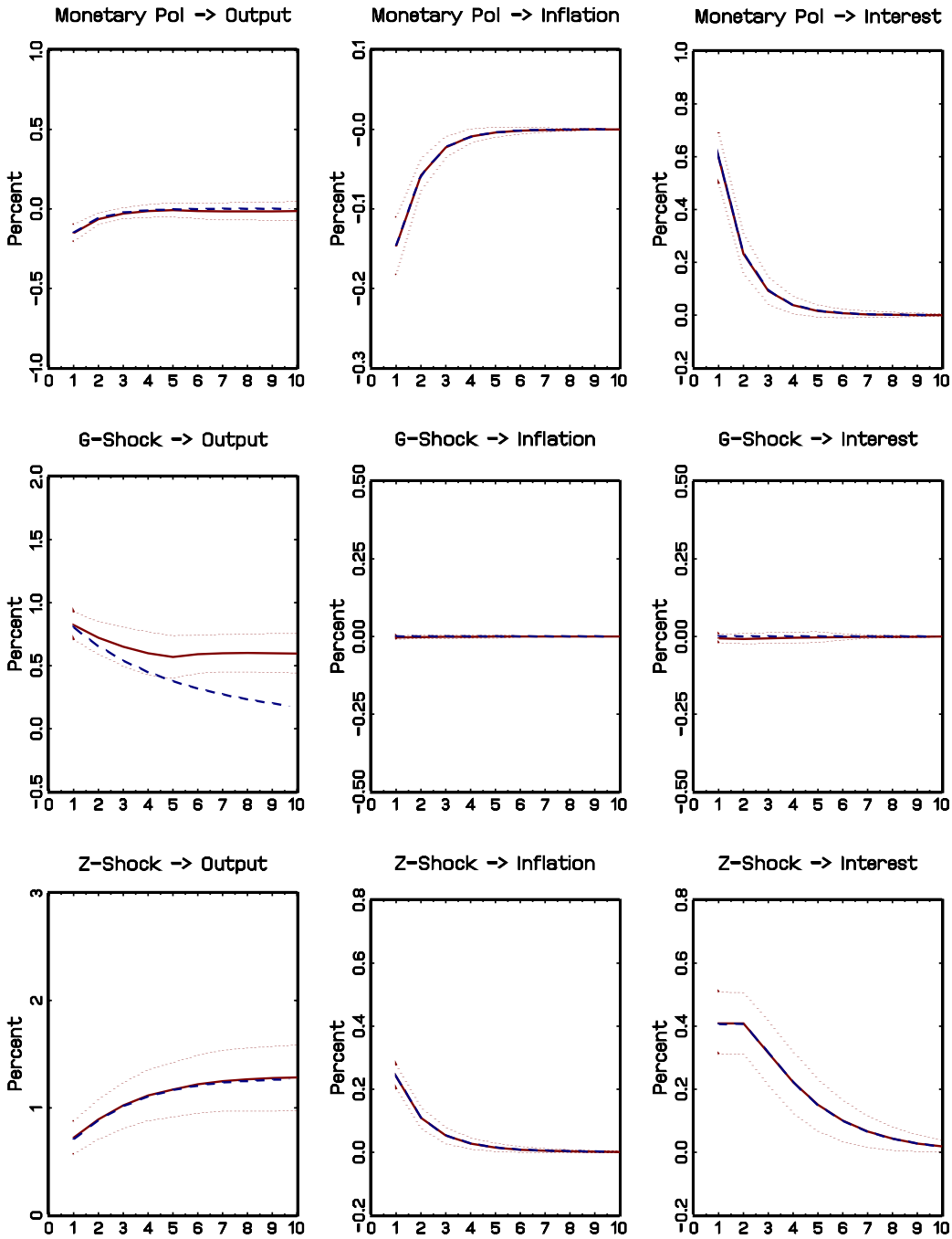
Notes: Output gap rule specification (top panel) and output growth rule specification (bottom panel), Data Sets 1-$\mathcal{M}_1$ and 1-$\mathcal{M}_2$, respectively. For each Markov chain, log marginal data densities are computed recursively with Geweke's modified harmonic mean estimator and plotted as a function of the number of draws.

Figure 10: Log Marginal Data Densities – Geweke vs. Chib-Jeliazkov



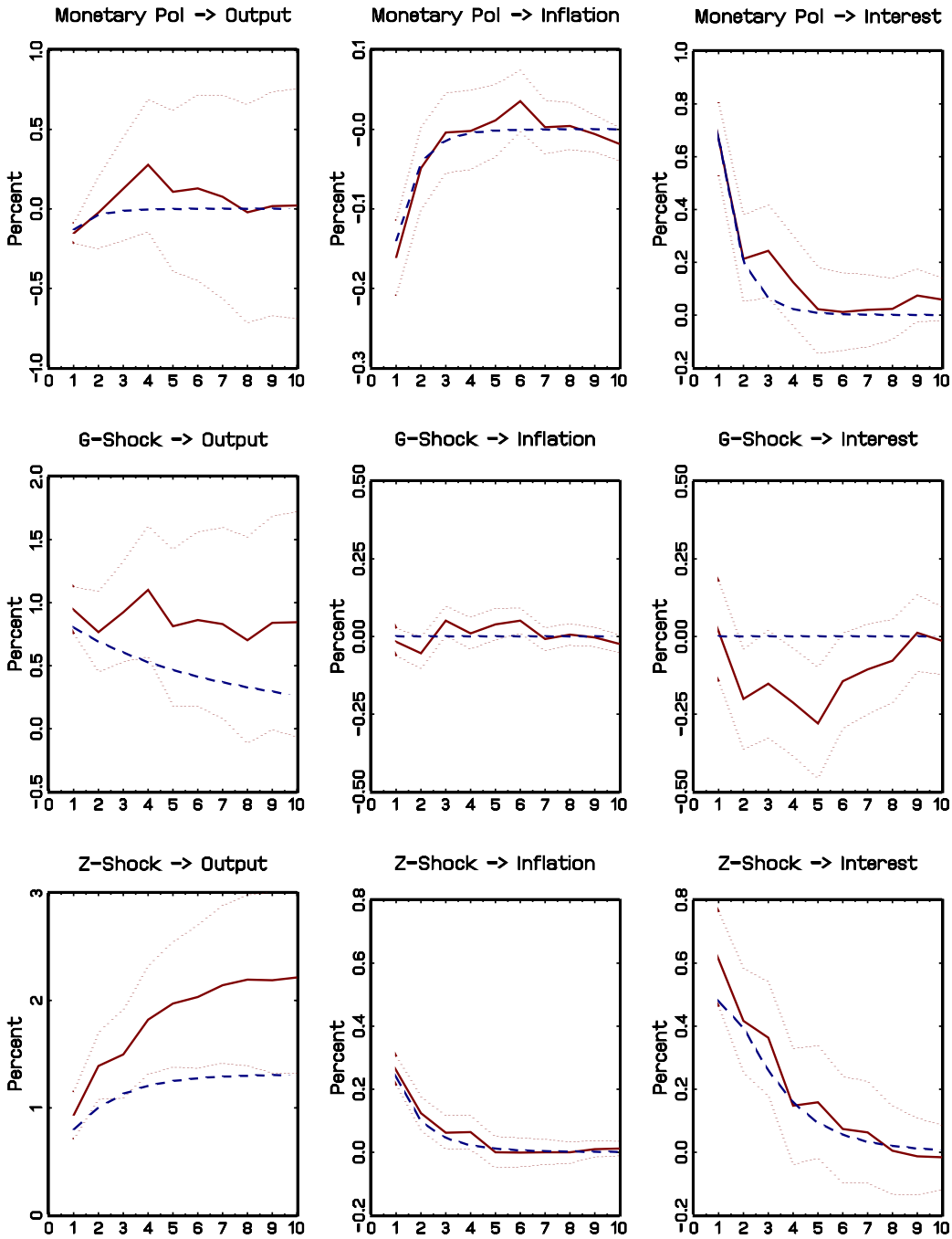
Notes: Output gap rule specification $\mathcal{M}_1$, Data Set 1-$\mathcal{M}_1$. Log marginal data densities are computed recursively with Geweke's modified harmonic mean estimator as well as the Chib-Jeliazkov estimator and plotted as a function of the number of draws.

Figure 11: Impulse Responses, DSGE and DSGE-VAR($\lambda = \infty$)



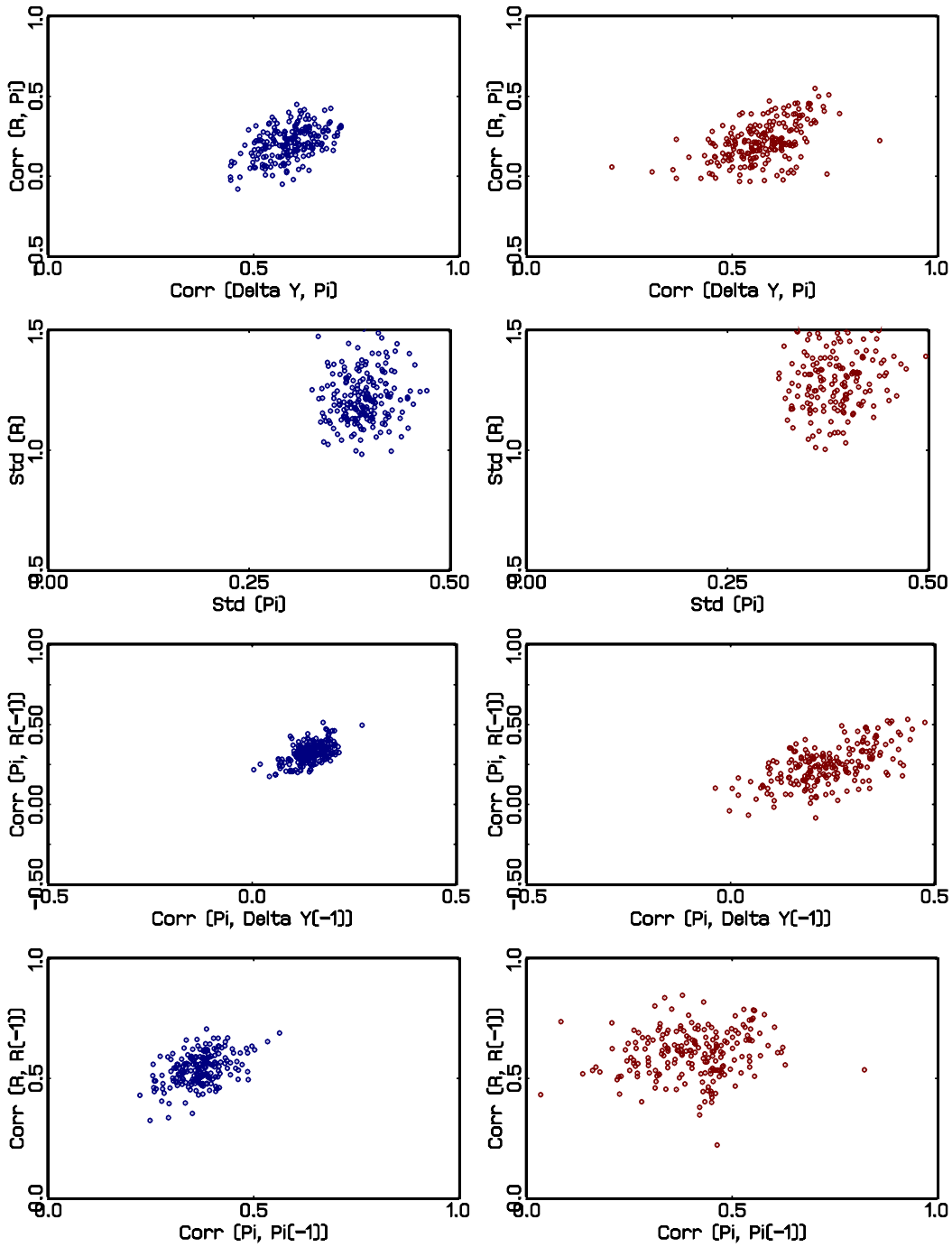
Notes: Output gap rule specification $\mathcal{M}_1$, Data Set 1-$\mathcal{M}_1$. DSGE model responses: posterior mean (solid); DSGE-VAR responses: posterior mean (dashed) and pointwise 90% probability bands (dotted).

Figure 12: Impulse Responses, DSGE and DSGE-VAR($\lambda = 0.75$)



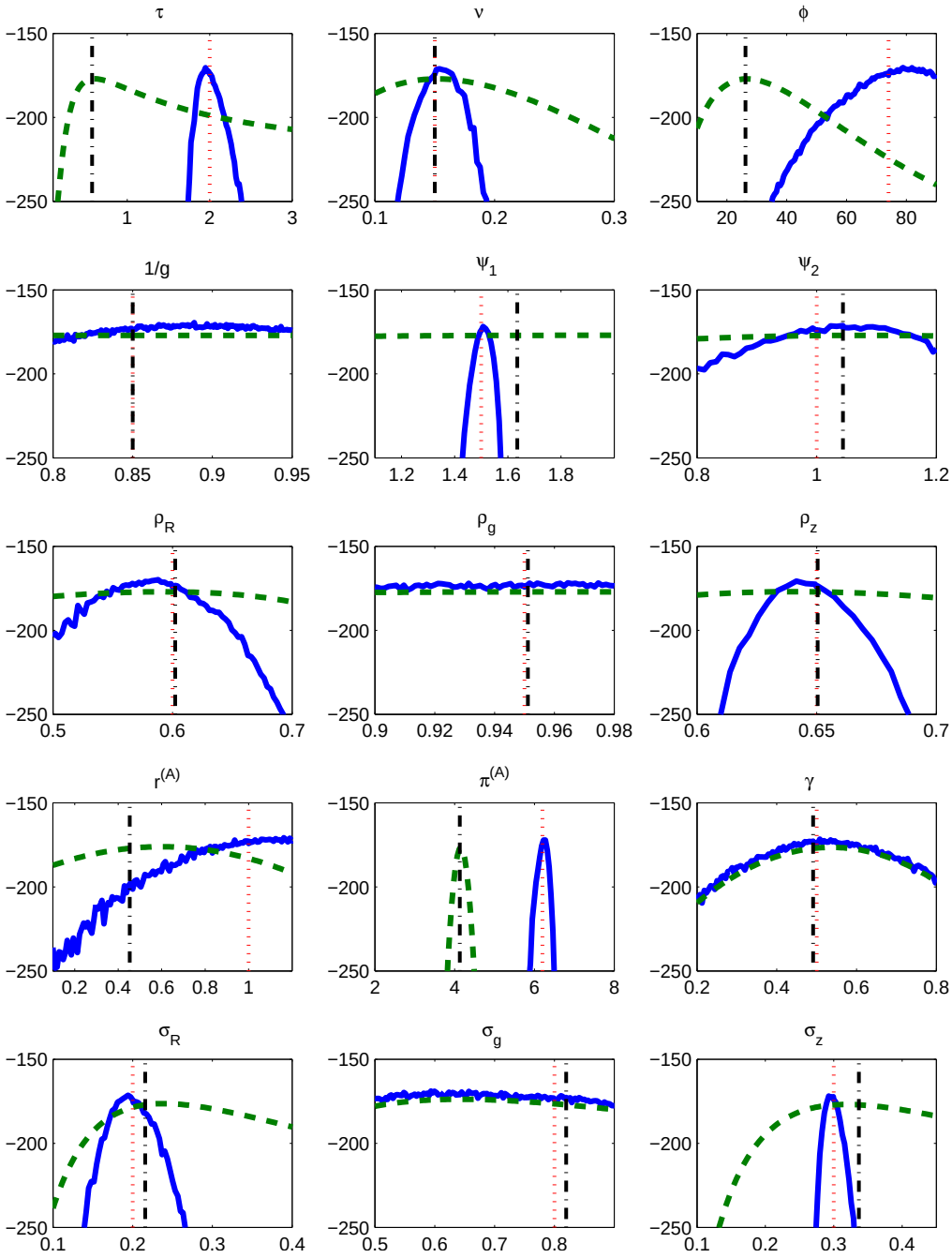
Notes: Output gap rule specification $\mathcal{M}_1$, Data Set 2. DSGE model responses: posterior mean (solid); DSGE-VAR responses: posterior mean (dashed) and pointwise 90% probability bands (dotted).

Figure 13: DSGE versus DSGE-VAR($\lambda = 0.75$) Posterior



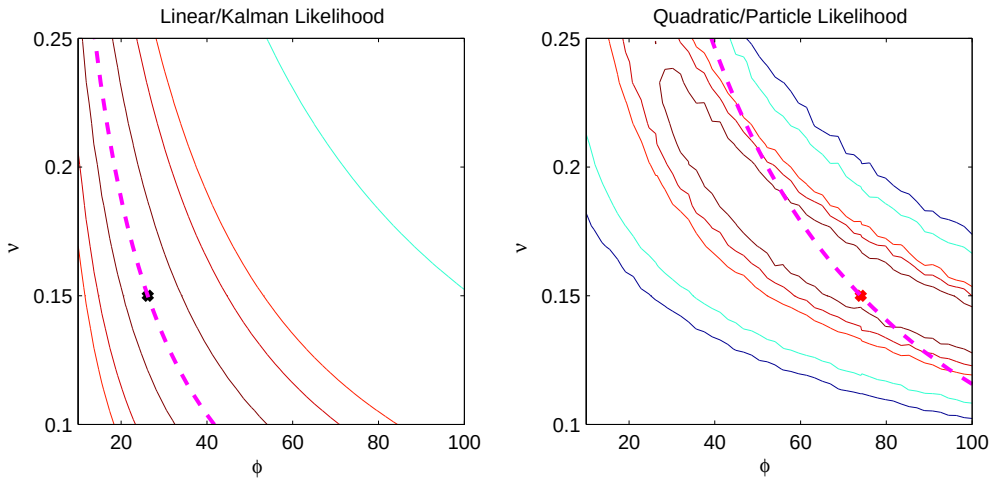
Notes: Output gap rule specification $\mathcal{M}_1$, Data Set 2. Panels depict 200 draws from the posterior distributions of selected population moments: DSGE model (left panels) and DSGE-VAR (right panels).

Figure 14: Linear vs. Quadratic Approximations: Likelihood Profiles



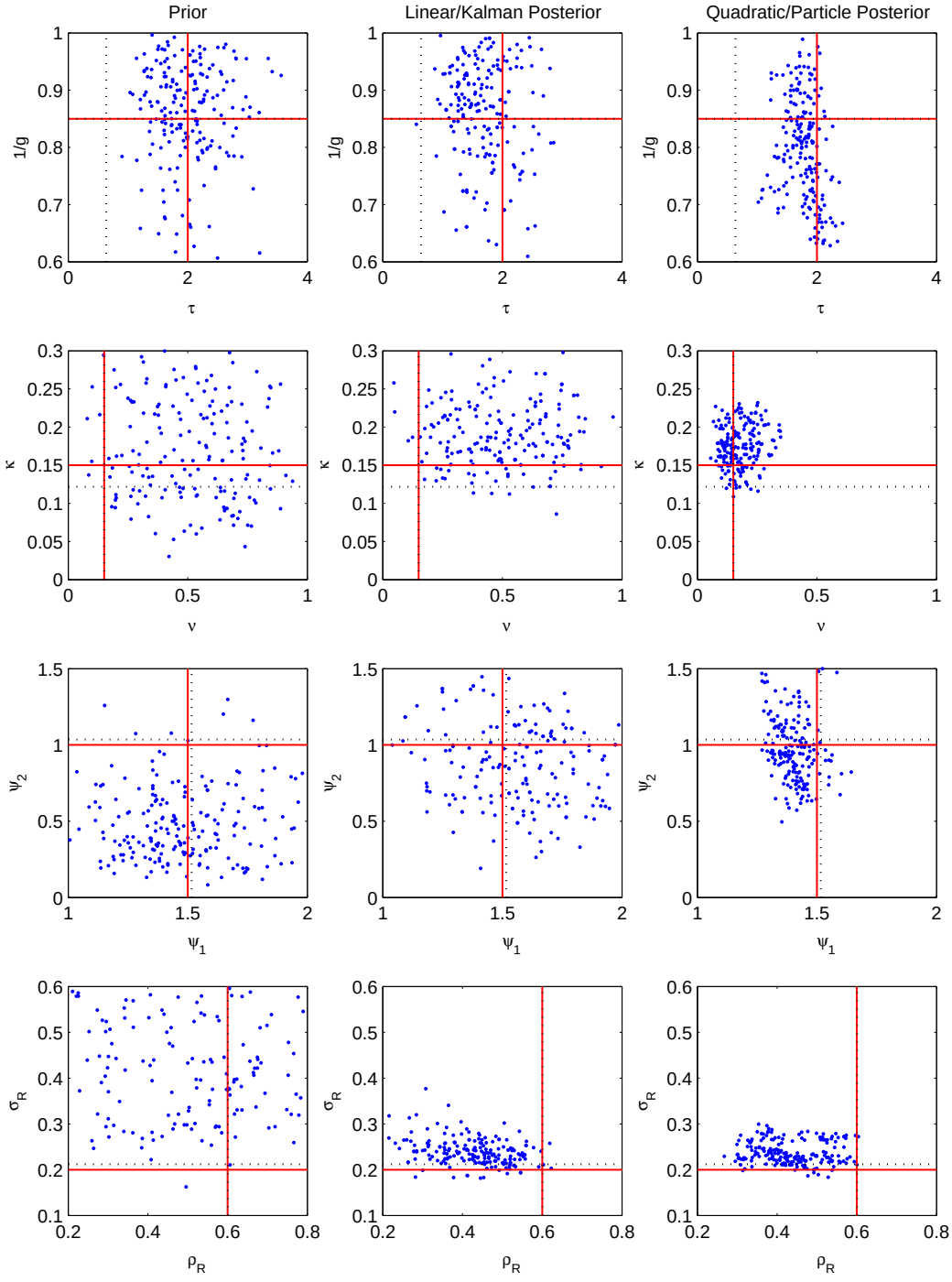
Notes: Output gap rule specification $\mathcal{M}_1$, Data Set 3. Likelihood profile for each parameter: linear/Kalman (dashed) and quadratic/particle (solid). 40,000 particles are used. True (dotted) and pseudo-true (dashdot) are shown.

Figure 15: Linear vs. Quadratic Approximations: Likelihood Contours



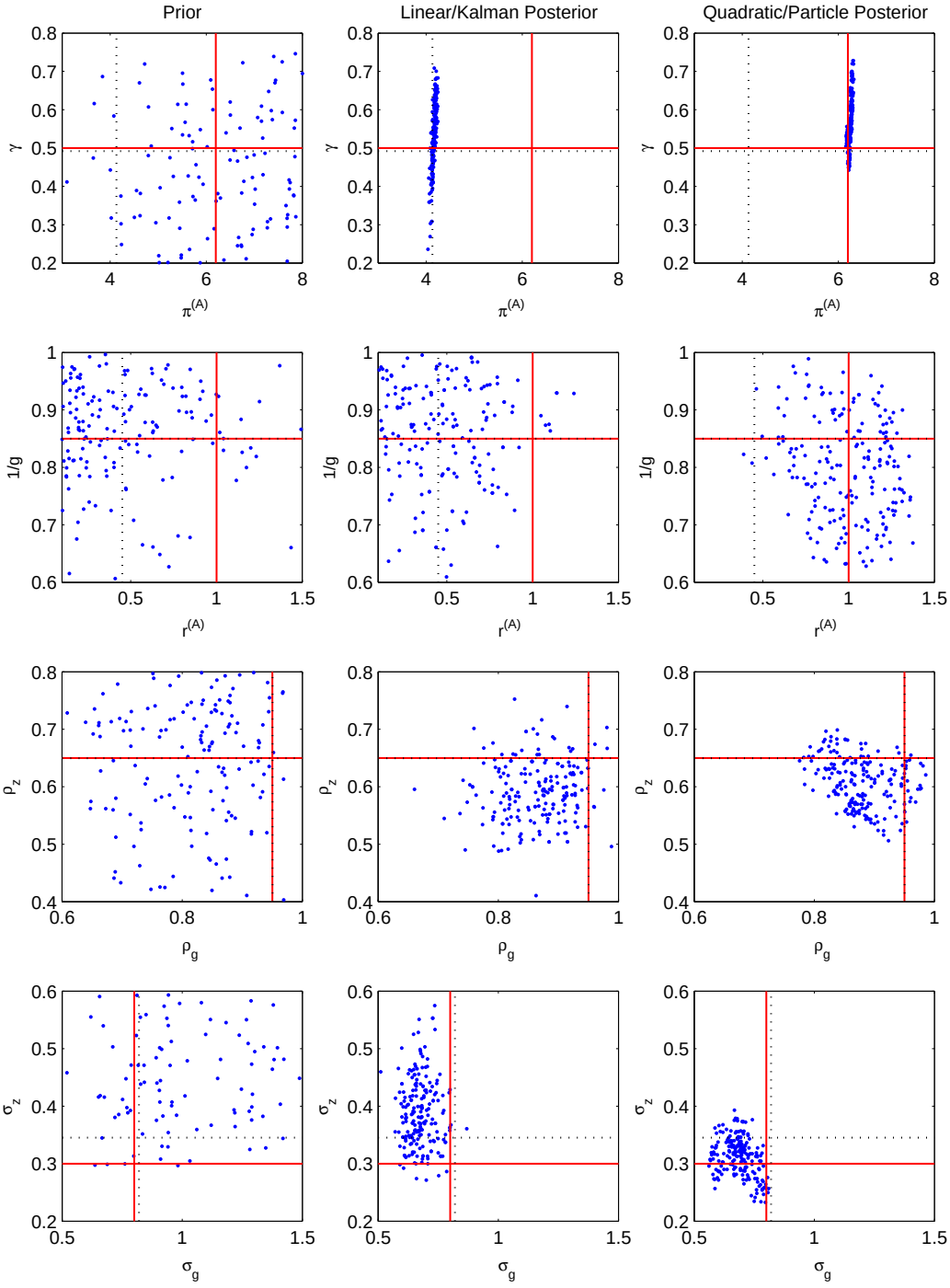
Notes: Output gap rule specification $\mathcal{M}_1$, Data Set 3. Contours of likelihood (solid). 40,000 particles are used. True and pseudo-true values are pointed. κ is constant along dashed line.

Figure 16: Posterior Draws: Linear versus Quadratic Approximation I



Notes: Output gap rule specification $\mathcal{M}_1$, Data Set 3. 100,000 draws from the prior and posterior distributions. Every 500th draw is plotted. True (solid) and pseudo-true (dotted) parameter values are shown. 40,000 particles are used.

Figure 17: Posterior Draws: Linear versus Quadratic Approximation II



Notes: Output gap rule specification $\mathcal{M}_1$, Data Set 3. 100,000 draws from the prior and posterior distributions. Every 500th draw is plotted. True (solid) and pseudo-true (dotted) parameter values are shown. 40,000 particles are used.