Quasi-Bayesian Model Selection[*]

Atsushi Inoue [†]   Mototsugu Shintani[‡]

Southern Methodist University   Vanderbilt University

December 2013

Abstract

In this paper we establish the consistency of the model selection criterion based on the quasi-marginal likelihood obtained from Laplace-type estimators (LTE). We consider cases in which parameters are strongly identified, weakly identified and partially identified. Our Monte Carlo results confirm our consistency results. Our proposed procedure is applied to select among monetary macroeconomic models using US data.

1 Introduction

Thanks to the development of fast computers and accessible software packages, Bayesian methods are now commonly used in the estimation of macroeconomic models. Bayesian estimators get around numerically intractable and ill-shaped likelihood functions, to which maximum likelihood estimators tend to succumb, by incorporating economically meaningful prior information. In a recent paper, Christiano, Trabandt and Walentin (2011) propose a new method of estimating a standard macroeconomic model based on the criterion function of the impulse response function matching estimator of Christiano, Eichenbaum and Evans (2005) combined with prior density. Instead of relying on a correctly specified likelihood function, they define an approximate likelihood function and proceed with a random walk Metropolis-Hastings algorithm. Chernozhukov and Hong (2003) establish that such an approach has a frequentist justification in a more general framework and call it a Laplace-type estimator (LTE) or quasi-Bayesian estimator.[1] Other applications of LTEs to estimate macroeconomic models include Christiano, Eichenbaum and Trabandt (2013) and Kormilitsina and Nekipelov (2013).

When two or more competing models are available, it is of great interest to select one model for policy analysis. When competing models are estimated by Bayesian methods, the models are often compared by their marginal likelihood. It is quite intuitive to compare models estimated by LTE using the "marginal likelihood" obtained from the LTE criterion function. In fact, Christiano, Eichenbaum and Trabandt (2013, Table 4) report the marginal likelihoods from LTE when they compare the performance of their macroeconomic model of wage bargaining with that of a standard labor search model. In this paper, we prove that such practice is asymptotically valid in that a model with a larger value of its marginal likelihood is either correct or a better approximation to true impulse responses with probability approaching one as the sample size goes to infinity.

We consider the consistency of model selection based on the marginal likelihood in three cases: (i) parameters are all strongly identified; (ii) some parameters are weakly

---

[1]We follow Chernozhukov and Hong (2003) and use the term "quasi-Bayesian" in our paper. However, it should be noted that the same term also refers to the procedure that involves data-dependent prior or multiple priors in the Bayesian literature.

identified; and (iii) some model parameters are partially identified. While case (i) is standard in the model selection literature (e.g., Phillips, 1996; Sin and White, 1996), cases (ii) and (iii) are also empirically relevant because some parameters may not be strongly identified in macroeconomic models (see Canova and Sala, 2009). We consider the case of weak identification using a device that is similar to Stock and Wright (2000) and Guerron-Quintana, Inoue and Kilian (2013). We also consider the case in which parameters are set identified as in Chernozhukov, Hong and Tamer (2007) and Moon and Schorfheide (2009).

Our approach allows for model misspecification and is similar in spirit to the Bayesian model selection procedure considered by Schorfheide (2000). Instead of using the marginal likelihoods (or the standard posterior odds ratio) directly, Schorfheide (2000) introduces the VAR model as a reference model in the computation of the loss function so that he can compare the performance of possibly misspecified dynamic stochastic general equilibrium (DSGE) models in the Bayesian framework. The related DSGE-VAR approach of Del Negro and Schorfheide (2004, 2009) also allows DSGE models to be misspecified, which results in a small weight on the DSGE model obtained by maximizing the marginal likelihood of the DSGE-VAR model. An advantage of our approach is that we can directly compare the (quasi-) marginal likelihoods even if all the competing DSGE models are misspecified.[2]

Using the frequentist framework, Corradi and Swanson (2007), Dridi, Guay and Renault (2007) and Hnatkovska, Marmer and Tang (2012), among others, propose the hypothesis testing procedures to evaluate the relative performance of possibly misspecified DSGE models. Among these studies, Hnatkovska, Marmer and Tang (2012) is most closely related to our approach in that their test statistic is based on the estimated objective function of impulse response matching, namely, the estimated distance between impulse responses of the estimated DSGE model and VAR model. While our method can be interpreted as a frequentist approach, it differs from their testing approach as we proposes a consistent model selection procedure.

The outline of this paper is as follows: Asymptotic justifications for the quasi-

---

[2]As established in White (1982), desired asymptotic results can be often obtained even if the likelihood function is misspecified. The quasi-Bayesian approach is also closely related to the limited-information likelihood principle used by Zellner (1998) and Kim (2002) among others.

marginal likelihood model selection criterion are established in Section 2. Computational issues are discussed in Section 3. A small set of Monte Carlo experiments is provided in Section 4. An empirical application is illustrated in Section 5. The concluding remarks are made in Section 6. All proofs are relegated to the appendix.

2   Asymptotic Theory

Let γ_A denote a $k_A \times 1$ vector of structural impulse responses obtained from a VAR model and $f(\alpha)$ be a $k_A \times 1$ vector of structural impulse responses implied by a DSGE model when the value of structural parameters is given by $\alpha \in A$ where $A \subset \Re^{p_A}$. The impulse response function matching estimator of Christiano, Eichenbaum and Evans (2005) minimizes the criterion function

$$\hat{q}_{A,T}(\alpha) \ = \ \frac{1}{2}(\hat{\gamma}_{A,T} - f(\alpha))'\widehat{W}_{A,T}(\hat{\gamma}_{A,T} - f(\alpha))$$

with respect to $\alpha \in A$, where $\hat{\gamma}_{A,T}$ is a $k_A \times 1$ vector of structural impulse responses obtained from an estimated VAR model, and $\widehat{W}_{A,T}$ is a $k_A \times k_A$ positive semidefinite weighting matrix.[3] Following Chernozhukov and Hong (2003), define the quasi-posterior by

$$\frac{(2\pi/T)^{-\frac{k_A}{2}}|\widehat{W}_{A,T}|^{\frac{1}{2}}e^{-T\hat{q}_{A,T}(\alpha)}\pi_A(\alpha)}{\int_A (2\pi/T)^{-\frac{k_A}{2}}|\widehat{W}_{A,T}|^{\frac{1}{2}}e^{-T\hat{q}_{A,T}(\alpha)}\pi_A(\alpha)d\alpha},$$

where $\pi_A(\alpha)$ is the prior probability density function. This Laplace-type estimator is particularly useful when the criterion function $\hat{q}_{A,T}(\alpha)$ is not numerically tractable or when extremum estimates are not reasonable.

We propose quasi-marginal likelihoods (QMLs) for selecting a model and establish the consistency of the model selection based on QMLs. Suppose that we compare DSGE models, models A and B. Models A and B are parameterized by structural parameter vectors, $\alpha \in A$ and $\beta \in B$, where $A \subset \Re^{p_A}$ and $B \subset \Re^{p_B}$, and imply vectors of structural impulse responses, $f(\alpha)$ and $g(\beta)$, of dimensions $k_A \times 1$ and $k_B \times 1$, respectively. While the competing models are typically estimated from the same set of

---

[3]Jórda and Kozicki (2011) develop a projection minimum distance estimator that is based on restrictions of the form of $h(\gamma, \alpha) = 0$. While we could consider a quasi-Bayesian estimator based on such restrictions, we focus on the special case in which $h(\gamma, \alpha) = \gamma - f(\alpha)$.

impulse responses in practice, we also allow for the case in which they are estimated from different sets of impulse responses.

Define the quasi-marginal likelihood for model A by

$$
\begin{aligned}
m_A &= (2\pi/T)^{-\frac{k_A}{2}} |\widehat{W}_{A,T}|^{\frac{1}{2}} \int_A e^{-\frac{T}{2}(\hat{\gamma}_{A,T}-f(\alpha))'\widehat{W}_{A,T}(\hat{\gamma}_{A,T}-f(\alpha))} \pi_A(\alpha)d\alpha \\
&= (2\pi/T)^{-\frac{k_A}{2}} |\widehat{W}_{A,T}|^{\frac{1}{2}} \int_A e^{-T\hat{q}_{A,T}(\alpha)} \pi_A(\alpha)d\alpha.
\end{aligned} \tag{1}
$$

Similarly, define the quasi-marginal likelihood for model B by

$$
\begin{aligned}
m_B &= (2\pi/T)^{-\frac{k_B}{2}} |\widehat{W}_{B,T}|^{\frac{1}{2}} \int_B e^{-\frac{T}{2}(\hat{\gamma}_{B,T}-g(\beta))'\widehat{W}_{B,T}(\hat{\gamma}_{B,T}-g(\beta))} \pi_B(\beta)d\beta \\
&= (2\pi/T)^{-\frac{k_B}{2}} |\widehat{W}_{B,T}|^{\frac{1}{2}} \int_B e^{-T\hat{q}_{B,T}(\beta)} \pi_B(\beta)d\beta,
\end{aligned} \tag{2}
$$

where $\hat{\gamma}_{B,T}$ is a $k_B \times 1$ vector of structural impulse responses obtained from an estimated VAR model and $\widehat{W}_{B,T}$ is a $k_B \times k_B$ positive semidefinite weighting matrix.

Let

$$
q_A(\alpha_0) = \frac{1}{2}(\gamma_{A,0} - f(\alpha_0))'W_A(\gamma_{A,0} - f(\alpha_0)), \tag{3}
$$

$$
q_B(\beta_0) = \frac{1}{2}(\gamma_{B,0} - g(\beta_0))'W_B(\gamma_{B,0} - f(\beta_0)), \tag{4}
$$

where $\gamma_{A,0}$ and $\gamma_{B,0}$ are vectors of population structural impulse responses, α_0 and β_0 are the (possibly pseudo) true parameter values of α and β, and W_A and W_B are positive definite matrices.

We say that the quasi-marginal likelihood model selection criterion is consistent if the following property holds: $m_A > m_B$ [$m_A < m_B$] with probability approaching one if $q_A(\alpha_0) < q_B(\beta_0)$ [$q_A(\alpha_0) > q_B(\beta_0)$].

2.1 The Case of Strongly Identified Parameters

First, consider the case in which the parameters are strongly identified, that is, there are unique parameter values of α and β that minimize the impulse response matching criterion functions in population, $q_A(\alpha)$ and $q_B(\beta)$. Note that strong identification does not necessarily imply that the models are correctly specified because it is possible that $q_A(\alpha_0) > 0$ or $q_B(\beta_0) > 0$. We will assume the following conditions:

Assumption 1

(a) A and B are compact in $\Re^{p_A}$ and $\Re^{p_B}$, respectively.

(b) There exist $\alpha_0 \in \text{int} A$ and $\beta_0 \in \text{int} B$ such that for every $\epsilon > 0$

$$\inf_{\alpha \in A : \|\alpha - \alpha_0\| \geq \epsilon} q_A(\alpha) > q_A(\alpha_0),$$
$$\inf_{\beta \in B : \|\beta - \beta_0\| \geq \epsilon} q_B(\beta) > q_B(\beta_0),$$

where $q_A(\alpha)$ and $q_B(\beta)$ are defined in (3) and (4), respectively.

(c) $\sup_{\alpha \in A} |\hat{q}_{A,T}(\alpha) - q_A(\alpha)| = o_p(1)$ and $\sup_{\beta \in B} |\hat{q}_{B,T}(\beta) - q_B(\beta)| = o_p(1)$, where

$$\hat{q}_{A,T}(\alpha) = \frac{1}{2}(\hat{\gamma}_{A,T} - f(\alpha))'\widehat{W}_{A,T}(\hat{\gamma}_{A,T} - f(\alpha)),$$
$$\hat{q}_{B,T}(\beta) = \frac{1}{2}(\hat{\gamma}_{B,T} - g(\beta))'\widehat{W}_{B,T}(\hat{\gamma}_{B,T} - g(\beta)).$$

(d) $f : A \to \Re^{k_A}$ and $g : B \to \Re^{k_B}$ are twice continuously differentiable in the interior of A and B, and their Jacobian matrices $\partial f(\alpha_0)/\partial \alpha'$ and $\partial g(\beta_0)/\partial \beta'$ have rank p_A and p_B, respectively.

(e) $\pi_A : A \to \Re_+$ and $\pi_B : B \to \Re_+$ are continuous in neighborhoods of α_0 and β_0 with $\pi_A(\alpha_0) > 0$ and $\pi_B(\beta_0) > 0$, respectively.

(f) $\widehat{W}_{A,T}$ and $\widehat{W}_{B,T}$ are positive semidefinite with probability one and converge in probability to positive definite matrices W_A and W_B, respectively.

Remarks

1. In this subsection we assume that the parameters are globally identified (Assumption 1b). Because W_A and W_B are positive definite (Assumption 1b) and the Jacobian matrices are of full rank (Assumption 1d), the Hessian matrices of $q_A(\alpha)$ and $q_B(\beta)$ are positive definite. Thus the parameters are also strongly identified under our assumptions.

2. Assumption 1c requires uniform convergence of $\hat{q}_{A,T}(\cdot)$ and $\hat{q}_{B,T}(\cdot)$ to $q_A(\cdot)$ and $q_B(\cdot)$, respectively, which holds under more primitive assumptions. For example, suppose that $\hat{\gamma}_{A,T} \xrightarrow{p} \gamma_{A,0}$ and $\widehat{W}_{A,T} \xrightarrow{p} W_A$. Then pointwise convergence of $\hat{q}_{A,T}(\alpha)$ to $q_A(\alpha)$ holds, i.e., $\hat{q}_{A,T}(\alpha) \xrightarrow{p} q_A(\alpha)$ for each $\alpha \in A$. Because f is continuously differentiable (Assumption 1d), it follows that $q_A(\alpha)$ is uniformly continuous and that $\hat{q}_{A,T}(\alpha)$ is a Lipschitz function. It follows from Lemma 1(a) of Andrews (1992, p. 246) that

$\hat{q}_{A,T}(\alpha)$ is stochastically equicontinuous. By Theorem 1 of Andrews (1992, p. 245), Assumption 1c follows from the compactness of the parameter spaces (Assumption 1a) and the stochastic equicontinuity.

3. Typical prior densities are continuous in macroeconomic applications, and Assumption 1(e) is likely to be satisfied.

To establish the consistency of quasi-marginal likelihood model selection criteria, it is useful to consider Laplace approximations of quasi-marginal likelihoods.

Lemma 1 (Validity of Laplace Approximations). Suppose that Assumption 1 holds: Then the following Laplace approximations to the quasi-marginal likelihoods hold:

$$m_A = e^{-T\hat{q}_{A,T}(\hat{\alpha}_T)}\left(\frac{T}{2\pi}\right)^{\frac{k_A-p_A}{2}}\pi_A(\hat{\alpha}_T)|\widehat{W}_{A,T}|^{\frac{1}{2}}|\nabla^2\hat{q}_{A,T}(\hat{\alpha}_T)|^{-\frac{1}{2}}(1+o_p(1)), \quad (5)$$

$$m_B = e^{-T\hat{q}_{B,T}(\hat{\beta}_T)}\left(\frac{T}{2\pi}\right)^{\frac{k_B-p_B}{2}}\pi_B(\hat{\beta}_T)|\widehat{W}_{B,T}|^{\frac{1}{2}}|\nabla^2\hat{q}_{B,T}(\hat{\beta}_T)|^{-\frac{1}{2}}(1+o_p(1)), \quad (6)$$

where $\hat{\alpha}_T$ and $\hat{\beta}_T$ are the classical minimum distance estimators of α and β, respectively, i.e., $\hat{\alpha}_T = \mathrm{argmin}_{\alpha\in A}\hat{q}_{A,T}(\alpha)$ and $\hat{\beta}_T = \mathrm{argmin}_{\beta\in B}\hat{q}_{B,T}(\beta)$.

While one could use Laplace approximations rather than estimate the quasi-marginal likelihood, it may not be feasible to compute Laplace approximations when identification is not strong. We use these approximations as a device for analyzing the asymptotic behavior of the quasi-marginal likelihood.

Hnatkovska, Marmer and Tang (2012) develop a Vuong-type quasi-likelihood ratio test for comparing macroeconomic models estimated by classical minimum distance estimators when $\hat{\gamma}_{A,T} = \hat{\gamma}_{B,T}$, $\gamma_{A,0} = \gamma_{B,0}$ and $k_A = k_B$. They consider cases of (i) nested, (ii) strictly non-nested and (iii) overlapping models. Let $\mathcal{F} = \{\gamma \in \Re^{k_A} : \gamma = f(\alpha) \text{ for some } \alpha \in A\}$ and $\mathcal{G} = \{\gamma \in \Re^{k_B} : \gamma = g(\beta) \text{ for some } \beta \in B\}$. Following Vuong's (1989) definition, we say that models A and B are nested if $\mathcal{F} \subset \mathcal{G}$ or $\mathcal{G} \subset \mathcal{F}$, strictly nonnested if $\mathcal{F} \cap \mathcal{G} = \emptyset$ and overlapping if they are neither nested or strictly nonnested. When the models are equally (in)correctly specified in terms of matching impulse responses, i.e., $q_A(\alpha_0) = q_B(\beta_0)$, Hnatkovska et al. (2012) show that $\hat{q}_{A,T}(\hat{\alpha}_T) - \hat{q}_{B,T}(\hat{\beta}_T) = O_p(T^{-1})$ if the models are nested and that $\hat{q}_{A,T}(\hat{\alpha}_T) - \hat{q}_{B,T}(\hat{\beta}_T) = O_p(T^{-\frac{1}{2}})$

7

if they are strictly nonnested or overlapping under some primitive assumptions.[4] First, we consider the case in which the models are nested.

Theorem 1 (Nested Models). Suppose that Assumption 1 holds.

(a) If $q_A(\alpha_0) < q_B(\beta_0)$ $[q_A(\alpha_0) > q_B(\beta_0)]$, then $m_A > m_B$ $[m_A < m_B]$ with probability approaching one.

(b) If $q_A(\alpha_0) = q_B(\beta_0)$ and $k_A - p_A > k_B - p_B$ $[k_A - p_A < k_B - p_B]$ and if $\hat{q}_{A,T}(\hat{\alpha}_T) - \hat{q}_{B,T}(\hat{\beta}_T) = O_p(T^{-1})$, then $m_A > m_B$ $[m_A < m_B]$ with probability approaching one.

Remarks

1. Theorem 1(a) shows that the proposed marginal likelihood model selection criterion selects the model with a smaller population impulse response matching criterion function with probability approaching one. Theorem 1(b) implies that, if the minimized population criterion functions take the same value, our model selection criterion will select the model with a greater number of overidentifying restrictions. This result is somewhat similar to that of Andrews' (1999) criterion, which selects as many correctly specified moment conditions as possible by including a bonus term that is increasing in the number of overidentifying restrictions. In the typical case in which the two models are estimated from the same set of impulse responses, i.e., $\gamma_{A,0} = \gamma_{B,0}$ and $k_A = k_B$, this result implies that the more parsimonious model will be chosen. In the special case where Model A is correctly specified and is a restricted version of Model B, our criterion will select Model A.

2. This consistency result applies whether or not the models are correctly specified or misspecified. If one model is correctly specified in that its minimized population criterion function is zero, while the other model is misspecified in that its minimized population criterion function is positive, our model selection criterion will select the correctly specified model with probability approaching one. Arguably, it may still make sense to minimize the criterion function even when two models are misspecified. Our model selection criterion will select the better approximating model with probability

---

[4]Technically, even in the overlapping case, if $f(\alpha_0) = g(\beta_0)$ we have $\hat{q}_{A,T}(\hat{\alpha}_T) - \hat{q}_{B,T}(\hat{\beta}_T) = O_p(T^{-1})$.

approaching one.

Next consider the case in which the two models are strictly nonnested or overlapping.

Theorem 2 (Strictly Nonnested or Overlapping Models). Suppose that Assumption 1 holds.

(a) If $q_A(\alpha_0) < q_B(\beta_0)$ $[q_A(\alpha_0) > q_B(\beta_0)]$, then $m_A > m_B$ $[m_A < m_B]$ with probability approaching one.

(b) If $q_A(\alpha_0) = q_B(\beta_0)$ and $T^{1/2}(\hat{q}_{A,T}(\hat{\alpha}_T) - \hat{q}_{B,T}(\hat{\beta}_T))$ converges in distribution to a nondegenerate zero-mean symmetric distribution, then $m_A > m_B$ with probability approaching one half.

Remarks. As in Theorem 1(a), Theorem 2(a) shows that the proposed marginal likelihood model selection criterion selects the model with a smaller population criterion function with probability approaching one. Unlike Theorem 1(b), however, the marginal likelihood does not necessarily select a more parsimonious model even asymptotically, when $q_A(\alpha_0) = q_B(\beta_0)$ and $f(\alpha_0) \neq g(\beta_0)$.

It is highly unlikely that different structural macroeconomic models match impulse responses equally well in the limit with the condition, $f(\alpha_0) \neq g(\beta_0)$. In such rare situations, however, one may still prefer a more parsimonious model based on Occam's razor or if a selected model is to be used for forecasting (Inoue and Kilian, 2006). For that purpose we propose the following *modified quasi-marginal likelihood*:

$$\tilde{m}_A = m_A e^{(T-\sqrt{T})\hat{q}_{A,T}(\hat{\alpha}_T)}, \tag{7}$$

$$\tilde{m}_B = m_B e^{(T-\sqrt{T})\hat{q}_{B,T}(\hat{\beta}_T)}. \tag{8}$$

The modified quasi-marginal likelihood effectively replaces $e^{-T\hat{q}_{A,T}(\hat{\alpha}_T)}$ in the Laplace approximation (5) by $e^{-\sqrt{T}\hat{q}_{A,T}(\hat{\alpha}_T)}$, and remains consistent for both nested and nonnested models. At the same time it selects a more parsimonious model even in the rare case in which $q_A(\alpha_0) = q_B(\beta_0)$ and $f(\alpha_0) \neq g(\beta_0)$.

Theorem 3 (Modified Quasi-Marginal Likelihood). Suppose that Assumption 1 holds.

(a) If $q_A(\alpha_0) < q_B(\beta_0)$ $[q_A(\alpha_0) > q_B(\beta_0)]$, then $\tilde{m}_A > \tilde{m}_B$ $[\tilde{m}_A < \tilde{m}_B]$ with probability approaching one.

(b) If $q_A(\alpha_0) = q_B(\beta_0)$ and $k_A - p_A > k_B - p_B$ $[k_A - p_A < k_B - p_B]$ and if $\hat{q}_{A,T}(\hat{\alpha}_T) - \hat{q}_{B,T}(\hat{\beta}_T) = O_p(T^{-\frac{1}{2}})$, then $\tilde{m}_A > \tilde{m}_B$ $[\tilde{m}_A < \tilde{m}_B]$ with probability approaching one.

Remarks. The modified quasi-marginal likelihood selects a parsimonious model if two models match impulse responses equally well, while it may have reduced power if one model matches impulse responses strictly better than the other. We will investigate this trade-off in Monte Carlo experiments.

2.2 The Case of Weakly Identified Parameters

It is well-known that some parameters of DSGE models may not be strongly identified. See Canova and Sala (2009) for examples of weak identification in DSGE models. It is therefore important to investigate asymptotic properties of our model selection procedure in case some parameters are weakly identified.

Following Guerron-Quintana, Inoue and Kilian (2013), we define weak identification in the minimum distance framework. Let $\alpha = [\alpha'_s \ \alpha'_w]'$, $A = A_s \times A_w$, $\beta = [\beta'_s \ \beta'_w]'$ and $B = B_s \times B_w$. We replace $f(\alpha)$ and $g(\beta)$ by

$$f_T(\alpha) = f_s(\alpha_s) + T^{-\frac{1}{2}} f_w(\alpha), \tag{9}$$

$$g_T(\beta) = g_s(\beta_s) + T^{-\frac{1}{2}} g_w(\beta), \tag{10}$$

respectively, where $f_s : A_s \to \Re^{k_A}$, $f_w : A \to \Re^{k_A}$, $g_s : B_s \to \Re^{k_B}$, $g_w : B \to \Re^{k_B}$, $A_s \subset \Re^{p_{A_s}}$ and $B_s \subset \Re^{p_{B_s}}$. Here α_s and β_s are strongly identified while α_w and β_w are weakly identified.

Assumption 2

(a) A and B are compact in $\Re^{p_A}$ and $\Re^{p_B}$, respectively.

(b) If $p_{A_s} > 0$ or $p_{B_s} > 0$ then there exist $\alpha_{s,0} \in \text{int} A_s$ and $\beta_{s,0} \in \text{int} B_s$ such that for every $\epsilon > 0$

$$\inf_{\alpha_s \in A_s : \|\alpha_s - \alpha_{s,0}\| \geq \epsilon} q_{A_s}(\alpha_s) > q_{A_s}(\alpha_{s,0}),$$
$$\inf_{\beta_s \in B_s : \|\beta_s - \beta_{s,0}\| \geq \epsilon} q_{B_s}(\beta_s) > q_{B_s}(\beta_{s,0}),$$

where

$$q_{A_s}(\alpha_s) = \frac{1}{2}(\gamma_{A,0} - f_s(\alpha_s))'W_A(\gamma_{A,0} - f_s(\alpha_s)),$$

$$q_{B_s}(\beta_s) = \frac{1}{2}(\gamma_{B,0} - g_s(\beta_s))'W_B(\gamma_{B,0} - g_s(\beta_s)).$$

(c) $\sup_{\alpha \in A} |\hat{q}_{A,T}(\alpha) - q_A(\alpha_s)| = o_p(1)$ and $\sup_{\beta \in B} |\hat{q}_{B,T}(\beta) - q_B(\beta_s)| = o_p(1)$.

(d) $f_s : A_s \to \Re^{k_A}$, $f_w : A \to \Re^{k_A}$, $g_s : B_s \to \Re^{k_B}$ and $g_w : B \to \Re^{k_B}$ are twice continuously differentiable in the interior of A and B, and their Jacobian matrices $\partial f_s(\alpha_{s,0})/\partial \alpha_s'$ and $\partial g_s(\beta_{s,0})/\partial \beta_s'$ have rank p_{A_s} and p_{B_s}, respectively.

(e)

$$F_s(\alpha_{s,0})'W_A F_s(\alpha_{s,0}) + [(\gamma_{A,0} - f_s(\alpha_{s,0}))'W_A \otimes I_{p_{A_s}}]\frac{\partial \text{vec}(F_s(\alpha_{s,0})')}{\partial \alpha_s'}$$

and

$$G_s(\beta_{s,0})'W_B G_s(\beta_{s,0}) + [(\gamma_{B,0} - g_s(\beta_{s,0}))'W_B \otimes I_{p_{B_s}}]\frac{\partial \text{vec}(G_s(\beta_{s,0})')}{\partial \beta_s'}$$

are nonsingular, where $F_s(\alpha_s) = \partial f_s(\alpha_s)/\partial \alpha_s'$ and $G_s(\beta_s) = \partial g_s(\beta_s)/\partial \beta_s'$.

(f) $\pi_A : A \to \Re_+$ and $\pi_B : B \to \Re_+$ are continuous in α_0 and β_0 with $\pi_A(\alpha_{s,0}, \alpha_w) > 0$ and $\pi_B(\beta_{s,0}, \beta_w) > 0$ on sets of positive measure in A_w and B_w, respectively.

(g) $\widehat{W}_{A,T}$ and $\widehat{W}_{B,T}$ are positive semidefinite with probability one and converge in probability to positive definite matrices W_A and W_B.

Remarks.

1. Assumptions 2(b)(c)(d)(f) for weakly identified parameters are almost identical to Assumptions 1(b)(c)(d)(e) for strongly identified parameters.

2. In the proof, we consider a profile estimator of strongly identified parameters given weakly identified parameters. Assumption 2(e) allows us to apply the implicit function theorem to write the profile estimator as a smooth function of weakly identified parameters.

3. Assumption 2(e) can be interpreted as a rank condition for local identification under possible misspecification. When the model is correctly specified, the assumption simplifies to the conventional assumption that the Jacobian matrix, $F_s(\alpha_{s,0})$, is of full rank.

Theorem 4 (Weak Identification). Suppose that Assumption 2 holds.

(a) If $q_{A_s}(\alpha_{s,0}) < q_{B_s}(\beta_{s,0})$ $[q_{A_s}(\alpha_{s,0}) > q_{B_s}(\beta_{s,0})]$, then $m_A > m_B$ and $\tilde{m}_A > \tilde{m}_B$ $[m_A < m_B$ and $\tilde{m}_A < \tilde{m}_B]$ with probability approaching one as $T \to \infty$.

(b) If $q_{A_s}(\alpha_{s,0}) = q_{B_s}(\beta_{s,0})$, $k_A - p_{A_s} > k_B - p_{B_s}$ $[k_A - p_{A_s} < k_B - p_{B_s}]$ and $\hat{q}_{A,T}(\hat{\alpha}_T) - \hat{q}_{B,T}(\hat{\beta}_T) = O_p(T^{-1})$, then $m_A > m_B$ and $\tilde{m}_A > \tilde{m}_B$ $[m_A < m_B$ and $\tilde{m}_A < \tilde{m}_B]$ with probability approaching one.

(c) If $q_{A_s}(\alpha_{s,0}) = q_{B_s}(\beta_{s,0})$, $k_A - p_{A_s} > k_B - p_{B_s}$ $[k_A - p_{A_s} < k_B - p_{B_s}]$ and $\hat{q}_{A,T}(\hat{\alpha}_T) - \hat{q}_{B,T}(\hat{\beta}_T) = O_p(T^{-1/2})$, then $\tilde{m}_A > \tilde{m}_B$ $[\tilde{m}_A < \tilde{m}_B]$ with probability approaching one.

Remarks. Part (a) shows that our criteria select the model with a smaller value of the population objective function. Part (b) shows that, when two nested models share the same population objective function value, our criteria select a model with a greater degree of overidentification (or a more parsimonious model if $k_A = k_B$). We show in part (c) that, when two strictly nonnested or overlapping models share the same population objective function value, only the modified marginal likelihood is useful for selecting a more parsimonious model. Parts (b) and (c) are similar to Theorems 2(b) and 3(b) except that the degree of parsimony is in terms of the number of strongly identified parameters.

2.3 The Case of Partially Identified Parameters

We say that the parameters are partially identified if

$$A_0 = \{\alpha_0 \in A : q_A(\alpha_0) = \min_{\alpha \in A} q_A(\alpha)\}$$

consists of more than one points (see Chernozhukov, Hong and Tamer, 2007). Moon and Schorfheide (2012) lists macroeconometric examples in which this type of identification arises. Similarly, we define

$$B_0 = \{\beta_0 \in B : q_B(\beta_0) = \min_{\beta \in B} q_B(\beta)\}.$$

Assumption 3

(a) A and B are compact in $\Re^{p_A}$ and $\Re^{p_B}$, respectively.

(b) There exist $A_0 \subset A$ and $B_0 \subset B$ such that, for every $\alpha_0 \in A_0$, $\beta_0 \in B_0$, $\epsilon > 0$

$$\inf_{\alpha \in (A_0^c)^{-\epsilon}} q_A(\alpha) > q_A(\alpha_0),$$
$$\inf_{\beta \in (A_0^c)^{-\epsilon}} q_B(\beta) > q_B(\beta_0),$$

where $q_A(\alpha)$ and $q_B(\beta)$ are defined in (3) and (4), respectively, and

$$(A_0^c)^{-\epsilon} = \{\alpha \in A : d(\alpha, A_0) \geq \epsilon\},$$
$$(B_0^c)^{-\epsilon} = \{\beta \in B : d(\beta, B_0) \geq \epsilon\}.$$

(c) $\sup_{\alpha \in A} |\hat{q}_{A,T}(\alpha) - q_A(\alpha)| = o_p(1)$ and $\sup_{\beta \in B} |\hat{q}_{B,T}(\beta) - q_B(\beta)| = o_p(1)$.

(d) $\int_{A_0} \pi_A(\alpha) d\alpha > 0$ and $\int_{B_0} \pi_B(\beta) d\beta > 0$.

(e) $\widehat{W}_T$ is positive semidefinite with probability one and converges in probability to a positive definite matrix W.

(f) $A_0 = \{\alpha_{s,0}\} \times A_{p,0}$ and $B_0 = \{\beta_{s,0}\} \times B_{p,0}$ if some parameters are strongly identified.

(g)

$$F_s(\alpha_{s,0})' W_A F_s(\alpha_{s,0}) + [(\gamma_{A,0} - f_s(\alpha_{s,0}))' W_A \otimes I_{p_{A_s}}] \frac{\partial \text{vec}(F_s(\alpha_{s,0})')}{\partial \alpha_s'}$$

and

$$G_s(\beta_{s,0})' W_B G_s(\beta_{s,0}) + [(\gamma_{B,0} - g_s(\beta_{s,0}))' W_B \otimes I_{p_{B_s}}] \frac{\partial \text{vec}(G_s(\beta_{s,0})')}{\partial \beta_s'}$$

are nonsingular, where $F_s(\alpha_s) = \partial f_s(\alpha_s)/\partial \alpha_s'$ and $G_s(\beta_s) = \partial g_s(\beta_s)/\partial \beta_s'$.

Remarks.

1. Assumptions 3(b) and (d) are a generalization of Assumptions 1(b) and (e), respectively, to sets.

2. When Assumption 3(f) holds, there are strongly identified parameters. Assumption 3(g) allows us to write a profile estimator of the strongly identified parameters as a smooth function of partially identified parameters.

Theorem 5 (Partial Identification).

(a) Suppose that Assumption 3(a)–(e) holds. If $\min_{\alpha \in A} q_A(\alpha) < \min_{\beta \in B} q_B(\beta)$ $[\min_{\alpha \in A} q_A(\alpha) > \min_{\beta \in B} q_B(\beta)]$, then $m_A > m_B$ and $\tilde{m}_A > \tilde{m}_B$ $[m_A < m_B$ and $\tilde{m}_A < \tilde{m}_B]$ with probability approaching one as $T \to \infty$.

(b) Suppose that Assumption 3 holds. If $q_{A_s}(\alpha_{s,0}) = q_{B_s}(\beta_{s,0})$, $k_A - p_{A_s} > k_B - p_{B_s}$ $[k_A - p_{A_s} < k_B - p_{B_s}]$ and $\hat{q}_{A,T}(\hat{\alpha}_T) - \hat{q}_{B,T}(\hat{\beta}_T) = O_p(T^{-1})$, then $m_A > m_B$ and $\tilde{m}_A > \tilde{m}_B$ $[m_A < m_B$ and $\tilde{m}_A < \tilde{m}_B]$ with probability approaching one.

(c) Suppose that Assumption 3 holds. If $q_{A_s}(\alpha_{s,0}) = q_{B_s}(\beta_{s,0})$, $k_A - p_{A_s} > k_B - p_{B_s}$ $[k_A - p_{A_s} < k_A - p_{B_s}]$ and $\hat{q}_{A,T}(\hat{\alpha}_T) - \hat{q}_{B,T}(\hat{\beta}_T) = O_p(T^{-1/2})$, then $\tilde{m}_A > \tilde{m}_B$ $[\tilde{m}_A < \tilde{m}_B]$ with probability approaching one.

Remark. Theorem 5(a) shows that even in the presence of partially identified parameters, our criteria select a model with a smaller value of the population estimation objective function. This result occurs because it is the value of the objective function, not the parameter value, that matters to model selection.

3   Computational Issues

In this section, we describe three methods for computing the quasi-marginal likelihood: Laplace approximations, Geweke's (1998) modified harmonic estimator and the estimator of Chib and Jeliazkov (2001). While these are standard methods for computing marginal likelihood in Bayesian analyses, we present these methods for practitioners who are interested in using the Laplace-type estimator.

To use the Laplace approximation, we evaluate

$$e^{-Tq_A(\hat{\alpha}_T)} \left(\frac{T}{2\pi}\right)^{\frac{k_A - p_A}{2}} |\widehat{W}_{A,T}|^{\frac{1}{2}} \pi_A(\hat{\alpha}_T) |\nabla^2 \hat{q}_{A,T}(\hat{\alpha}_T)|^{-\frac{1}{2}}, \tag{11}$$

at the quasi-posterior mode, $\hat{\alpha}_T$. In our Monte Carlo experiment, we use 20 randomly chosen starting values for a numerical optimization routine to obtain the posterior mode. The weighting matrix can be either the diagonal matrix whose diagonal elements are the reciprocal of the bootstrap variances of impulse responses or the inverse of the bootstrap covariance matrix of impulse responses. In classical minimum distance estimation of impulse response matching, it is quite common to use the diagonal weighting matrix because estimators based on the optimal weighting matrix can be often economically implausible. We use 1,000 bootstrap replications to obtain the bootstrap covariance matrix of impulse response function estimators in the Monte Carlo experiments.

For the modified harmonic mean estimator and the Chib-Jeliazkov estimator, we follow the random walk Metropolis-Hastings algorithm in An and Schorfheide (2007). The proposal distribution is $N(\alpha^{(j-1)}, c\hat{H}^{-1})$ where $\alpha^{(0)} = \hat{\alpha}_T$, $c = 0.3$ for $j = 1$, $c = 1$ for $j > 1$ and $\hat{H}$ is the Hessian of the log-quasi-posterior evaluated at the quasi-posterior mode.[5] The draw α from $N(\alpha^{(j-1)}, c[\nabla^2 \hat{q}_{A,T}(\hat{\alpha}_T)]^{-1})$ is accepted with probability

$$\min\left(1, \frac{e^{-T\hat{q}_{A,T}(\alpha)}\pi_A(\alpha)}{e^{-T\hat{q}_{A,T}(\alpha^{(j-1)})}\pi_A(\alpha^{(j-1)})}\right). \tag{12}$$

In each Monte Carlo iteration, we use the second half of 100,000 draws to estimate the marginal likelihood. To satisfy the generalized information equality of Chernozhukov and Hong (2003), which is necessary for the validity of the MCMC method for the Laplace-type estimator, we need to use the inverse of the bootstrap covariance matrix of impulse response function estimators as $\widehat{W}_{A,T}$. Thus, the optimal weighting matrix is used for the modified harmonic mean estimator and the estimator of Chib and Jeliazkov (2001) because they are calculated from MCMC draws.

For the modified harmonic mean estimator, we first evaluate

$$E(e^{-T\hat{q}_T(\alpha)}\pi_A(\alpha)w(\alpha)) \tag{13}$$

by the MCMC draws, where

$$
\begin{aligned}
w(\alpha) &= \frac{1}{(1-p)} \frac{1}{((2\pi)^{p_A}|\widehat{Cov}(\alpha)|)^{\frac{1}{2}}} \exp\left(-\frac{(\alpha - \tilde{\alpha}_T)'[\widehat{Cov}(\alpha)]^{-1}(\alpha - \tilde{\alpha}_T)}{2}\right) \\
&\quad \times 1((\alpha - \tilde{\alpha}_T)'[\widehat{Cov}(\alpha)]^{-1}(\alpha - \tilde{\alpha}_T) < \chi^2_{p_A, 1-p}),
\end{aligned} \tag{14}
$$

$\tilde{\alpha}_T$ is the quasi-posterior mean, $\widehat{Cov}(\alpha)$ is the quasi-posterior covariance matrix, $\chi^2_{p_A, 1-p}$ is the $100(1-p)$ percentile of the chi-square distribution with p_A degrees of freedom. The modified harmonic mean estimator is the reciprocal of (13). In our Monte Carlo experiment p is set to 0.10.

For the estimator of Chib and Jeliazkov (2001), estimate the log of the quasi-marginal likelihood by

$$\ln \pi_A(\tilde{\alpha}_T) - T\hat{q}_{A,T}(\tilde{\alpha}_T) - \ln \hat{p}_A(\tilde{\alpha}_T) \tag{15}$$

where

$$\hat{p}_A(\alpha) = \frac{(1/J)\sum_{j=1}^{J} r(\alpha^{(j)}, \tilde{\alpha})\phi_{\tilde{\alpha}, c^2\tilde{\Sigma}}(\alpha^{(j)})}{(1/K)\sum_{k=1}^{K} r(\tilde{\alpha}, \alpha^{(k)})} \tag{16}$$

---

[5]When the parameters are partially identified, we set $c = 0.001$ to increase the acceptance rate.

where the numerator is evaluated using the second half of MCMC draws and the denominator is evaluated using $\alpha^{(k)}$ from $N(\tilde{\alpha}, c^2\tilde{\Sigma})$. In our Monte Carlo experiment, K is set to 50,000. $c^2\tilde{\Sigma}$ is either set to the one used in the proposal density or estimated from the posterior draws.

To compute the modified quasi-marginal likelihood, we use

$$\tilde{m}_A = m_A e^{(T-\sqrt{T})\hat{q}_{A,T}(\hat{\alpha}_T)}, \tag{17}$$

where m_A is the Laplace approximation, the modified harmonic mean estimator or the Chib-Jeliazkov estimator.

4   Monte Carlo Experiments

We conduct a simple Monte Carlo simulation for the purpose of investigating theoretical predictions obtained in the previous sections. Here, we employ a simple model based on one of the models considered by Canova and Sala (2009):

$$y_t = E_t(y_{t+1}) - \sigma(i_t - E_t(\pi_{t+1})) + u_{1t}, \tag{18}$$

$$\pi_t = E_t(\pi_{t+1}) + \kappa y_t + u_{2t}, \tag{19}$$

$$i_t = E_t(\pi_{t+1}) + u_{3t}, \tag{20}$$

where u_{1t}, u_{2t}, u_{3t} are independent iid standard normal random variables. Because the solution is

$$
\begin{bmatrix} y_t \\ \pi_t \\ i_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & -\sigma \\ \kappa & 1 & -\sigma\kappa \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}, \tag{21}
$$

we have covariance restrictions:

$$
Cov([y_t\ \pi_t\ i_t]') = \begin{bmatrix} 1+\sigma^2 & \kappa + \kappa\sigma^2 & -\sigma \\ \kappa + \sigma^2\kappa & 1+\kappa^2+\sigma^2\kappa^2 & -\sigma\kappa \\ -\sigma & -\sigma\kappa & 1 \end{bmatrix}. \tag{22}
$$

We consider cases of strong identification, weak identification and partial identification as well as cases of $q_A(\alpha_0) < q_B(\beta_0)$ and $q_A(\alpha_0) = q_B(\beta_0)$. In the first two designs, design 1 and design 2, we use

$$f(\sigma, \kappa) = [1+\sigma^2,\ \kappa + \sigma^2\kappa,\ -\sigma,\ 1+\kappa^2+\sigma^2\kappa^2,\ -\sigma\kappa]', \tag{23}$$

and the corresponding five elements in the covariance matrix of the three observed variables, where we set $\sigma = 1$ and $\kappa = 0.5$. In these two designs, the parameters are globally and locally identified. In design 1, the two parameters are estimated in model A, while σ is estimated and the value of κ is set to a wrong parameter value, 1, in model B. In other words, model A is correctly specified and model B is incorrectly specified. In design 2, only one parameter (σ) is estimated and the value of κ is set to the true parameter value in model A, while the two parameters are estimated in model B. While the two models are both correctly specified in this design, model A is more parsimonious than model B.

In the next two designs, design 3 and design 4, we use

$$f(\sigma, \kappa) \ = \ [\kappa + \sigma^2 \kappa, \ \ 1 + \kappa^2 + \sigma^2 \kappa^2, \ \ -\sigma \kappa]' \tag{24}$$

and the corresponding three elements of the covariance matrix are used. As κ approaches zero, the strength of identification of σ becomes weaker. We set $\sigma = 1$ and $\kappa = 0.5$. Designs 3 and 4 correspond to designs 1 and 2. In design 3, model B is incorrectly specified in that κ is set to 1. In design 4, the two models are both correctly specified and model A is more parsimonious than model B.

In the last two design, designs 5 and 6, the parameters are partially identified in that we estimate α and ζ and the restrictions depend on them only through $\kappa = (1 - \alpha)(1 - 0.99\alpha)\zeta/\alpha$. We use the five restrictions used in designs 1 and 2, (23), and we set $\sigma = 1$, $\alpha = 0.5$ and $\zeta = 1$ so that $\kappa \approx 0.5$ as in design 1. In design 5, two parameters, α and ζ, are estimated in model A, while the value of σ is set to the correct value, 1, in model A and is set to an incorrect value, 0.5, in model B. In design 6, only α and ζ are estimated while the value of σ is set to the true value in model A, whereas the three parameters are all estimated in model B. Note that in each of the six designs, model A is always preferred to model B because model A is correctly specified in designs 1, 3 and 5 and is more parsimonious in designs 2, 4 and 6. Table 1 summarizes the six designs.

The number of Monte Carlo replications is set to 1,000, the number of randomly chosen initial values for numerical optimization is set to 20, and the number of Markov Chain Monte Carlo draws is set to 100,000. The flat prior is used for all the parameters. The sample sizes are 50, 100 and 200.

We estimate the marginal likelihood of each model by ten methods: the Laplace approximation with the diagonal weighting matrix, the Laplace approximation with the optimal weighting matrix, the modified harmonic mean estimator, the estimator of Chib and Jeliazkov with the analytical covariance matrix of the proposal density used in the numerator of (16), their estimator with the covariance matrix estimated from the posterior draws used in the numerator of (16) and the modified marginal likelihood for each of these five estimators.

The optimal weighting matrix is used for the modified harmonic mean estimator, the Chib and Jeliazkov estimators and their modified versions. The details of these methods can be found in the preceding section. In designs 5 and 6, where the two parameters are only partially identified, we find that the Hessian of the log-posterior is never positive definite at the quasi-posterior mode and we do not use a Laplace approximation because it is infeasible. In designs 5 and 6, we add $2^{-26}I$ to the Hessian of the log-posterior to make it positive definite.

Table 2 reports the probabilities of selecting model A when the parameters are strongly identified. In design 1, our proposed criteria tend to select the correctly specified model (model A) over the incorrectly specified model (model B) regardless of the methods used to estimate the quasi-marginal likelihood. Because the modification halves the divergence rate of the quasi-marginal likelihood, the frequencies of selecting model A based on the modified quasi-marginal likelihood is not as high as those based on the quasi-marginal likelihood.

In design 2, the frequencies of choosing model A based on the quasi-marginal likelihood are not as high as those in design 1 because the consistency depends on logarithmic rates as opposed to linear rates (in terms of the log quasi-marginal likelihood). As the sample size grows, however, the probabilities approach one as expected. The modified quasi-marginal likelihood performs better in this design.

Table 3 shows that the probability of selecting model A also tends to approach one as the sample size increases, even when identification is weak (designs 3 and 4). In design 3, the modified quasi marginal likelihood inherits its performance in design 1 and does not perform as well as the quasi marginal likelihood. Especially, the modified quasi marginal likelihood based on the diagonal weighting matrix performs poorly in design 3 and may require much larger sample sizes.

18

When some parameters are partially identified, the methods also tend to select model A with probability approaching one as the sample size grows. The Chib-Jeliazkov estimator with estimated covariance matrix performs better than the one with analytical covariance matrix in design 6, and their performances are similar in design 5. The modified quasi marginal likelihood does not perform well in design 5. This result may be due to the poor performance of numerical optimization due to partial identification which results in inaccurate modification factors for these criteria.

Overall, we find that the modified harmonic mean and Chib-Jeliazkov estimators are less sensitive to properties of the Hessian of the log-posterior and perform well. Our proposed modified quasi-marginal likelihood is less powerful than the quasi-marginal likelihood when comparing a correct model and an incorrect model because of the halved divergence rate. (cases considered in part (a) of Theorems 1–5).

5 Empirical Application

In this section, we apply our procedure to a more empirically relevant medium-sized DSGE model originally developed by Christiano, Eichenbaum and Evans (2005) (hereafter CEE). In particular, we consider a modified version of the CEE model, which has been estimated by Smets and Wouters (2007), Altig, Christiano, Eichenbaum and Linde (2011), Christiano, Trabandt and Walentin (2011), and Christiano, Eichenbaum and Trabandt (2013), among others. The model is one of the most commonly used macroeconomic models among practitioners that incorporates the investment adjustment cost, habit formation in consumption, sticky prices and wages and the inflation-targeting monetary policy. In practice, this class of the model has been estimated using various methods. For example, CEE and Altig, Christiano, Eichenbaum and Linde (2011) employ the classical impulse response matching estimator, while Smets and Wouters (2007) estimate the model using a standard Bayesian method. As a third approach, Christiano, Trabandt and Walentin (2011), and Christiano, Eichenbaum and Trabandt (2013) employ the quasi-Bayesian impulse response matching estimator, or the Laplace-type estimator.

For the purpose of evaluating the relative importance of various frictions in the model estimated by the standard Bayesian method, Smets and Wouters (2007) uti-

lize the marginal likelihood. Their question is whether all the frictions introduced in the canonical DSGE model are really necessary in order to describe the dynamics of observed aggregate data. To answer this question, they compare marginal likelihoods of estimated models when each of the frictions was drastically reduced one at time. Among the sources of nominal frictions, they claim that both price and wage stickiness are equally important while indexation is relatively unimportant in both goods and labor markets. Regarding the real frictions, they claim that the investment adjustment costs are most important. They also find that, in the presence of wage stickiness, the introduction of variable capacity utilization is less important.

Here, we conduct a similar exercise using quasi-marginal likelihoods (QMLs) obtained in the quasi-Bayesian impulse response matching estimation. The data and estimated impulse response functions are identical to the ones used in Christiano, Trabandt and Walentin (2011).[6] They estimate a VAR(2) model of 14 variables using the US quarterly data from 1951Q1 to 2008Q4. Then, a combination of short-run and long-run restrictions is used to identify the responses to three types of shocks in the economy: (i) a monetary policy shock, (ii) a neutral technology shock and (iii) an investment-specific technology shock. All the specifications of shock processes we employ here are same as those used in Christiano, Trabandt and Walentin (2011). In respect to the monetary policy shock, the interest rate R_t is assumed to follow the process given by

$$\ln(R_t/R) = \rho_R \ln(R_{t-1}/R) + (1 - \rho_R)\left[r_\pi \ln(\pi_{t+1}/\pi) + r_y \ln(gdp_t/gdp)\right] + \varepsilon_{R,t}$$

where gdp_t is scaled real GDP and $\varepsilon_{R,t} \sim iid(0, \sigma_R^2)$. The neutral technology Z_t in log is assumed to be I(1) with its growth generated from an iid process

$$g_{Z,t} = \gamma_Z + \varepsilon_{Z,t}$$

where $g_{Z,t} = \ln(Z_t/Z_{t-1})$ and $\varepsilon_{Z,t} \sim iid(0, \sigma_Z^2)$. The investment-specific technology Ψ_t in log is also assumed to be I(1) but its growth is generated from an AR(1) process given by

$$g_{\Psi,t} = (1 - \rho_\Psi)\gamma_\Psi + \rho_\Psi g_{\Psi,t-1} + \varepsilon_{\Psi,t}$$

---

[6]See the data appendix of their paper for the detailed explanation.

where $g_{\Psi,t} = \ln(\Psi_t/\Psi_{t-1})$ and $\varepsilon_{\Psi,t} \sim iid(0, \sigma_\Psi^2)$. The structural parameters are estimated by matching the first 15 responses of selected 9 variables to 3 shocks, less 8 zero contemporaneous responses to the monetary policy shock (so that the total number of responses to match is 397).

Since our purpose is to evaluate the relative contribution of various frictions, we estimate some additional parameters, such as the wage stickiness parameter ξ_w, wage indexation parameter ι_w and price indexation parameter ι_p, which are fixed in the analysis of Christiano, Trabandt and Walentin (2011).[7] The list of estimated structural parameters in our analysis, quasi-Bayesian estimates and the prior distribution, are reported in Table 5. The estimated impulse response functions are provided in Figures 1, 2 and 3. This estimated model serves as the baseline model when we compare with other models using QMLs.

We follow Smets and Wouters (2007) and divide the sources of frictions of the baseline model into two groups. First, nominal frictions are sticky prices, sticky wages, price indexation and wage indexation. Second, real frictions are investment adjustment costs, habit formation, capital utilization. We estimate additional submodels, which reduces the degree of each of the seven frictions. The computed QMLs for 8 models, including the baseline model, are reported in Table 6. Both QMLs based on the Laplace approximation and the modified harmonic mean estimator are reported. For the reference, also included in the table are the original marginal likelihoods obtained by Smets and Wouters (2007) based on the different estimation method applied to the different data set. The first column shows the quasi-posterior mean of relevant structural parameters of the baseline model along with the QML.

Let us first examine the relative role of nominal frictions. The second and third columns show the results when the degree of nominal price and wage stickiness (ξ_p and ξ_w) is set at 0.10, respectively. Consistent with Smets and Wouters (2007), the results show the importance of both types of nominal frictions. Unlike the result obtained by Smets and Wouters (2007), however, the QML deteriorates more by restricting the degree of wage stickiness. The fourth and fifth columns show the results when the price and wage indexation parameters (ι_p and ι_w) are set at 0.01, respectively. Again, consistent with Smets and Wouters (2007), neither price nor wage indexation plays a

---

[7]In our analysis, both price markup and wage markup parameters are fixed at 1.2.

very important role in terms of improving the value of QMLs. The value of QML is similar to that of baseline model even when the price indexation parameter is restricted to a very low value. In fact, when the wage indexation parameter is set at a small value, QML seems to improve over the baseline model. Thus, we can conclude that Calvo-type frictions in price and wage settings are empirically more important than the price and wage indexation to past inflation. Let us now turn to the role of real frictions. The remaining three columns show the results when each of investment adjustment cost parameter (S''), consumption habit parameter (b) and capital utilization cost parameter (σ_a) is set at some small values. The results show that restricting habit formation in consumption significantly reduces the QML compared to other two real frictions, suggesting the relatively important role of the consumption habit. Our result on the role of capital utilization costs is also somewhat similar to the one obtained by Smets and Wouters (2007) in the sense that it has a relatively minor role in increasing the fit of the model. Overall, our results seem to support the empirical evidence obtained by Smets and Wouters (2007), despite the fact that our analysis is based on a very different model selection criterion.

6   Concluding Remarks

In this paper we established the consistency of the model selection criterion based on the quasi-marginal likelihood obtained from Laplace-type estimators (LTE). We considered cases in which parameters are strongly identified and are weakly identified. Our Monte Carlo results confirmed our consistency results. Our proposed procedure was also applied to select an appropriate specification in monetary macroeconomic models using US data.

Appendix

Proof of Lemma 1.

Let $\hat{\alpha}_T = \operatorname{argmin}_{\alpha \in A} \hat{q}_{A,T}(\alpha)$, and $\hat{\beta}_T = \operatorname{argmin}_{\beta \in B} \hat{q}_{B,T}(\beta)$. Under Assumptions

1(a)–(c), $\hat{\alpha}_T$ and $\hat{\beta}_T$ uniquely exist with probability approaching one. It follows from Theorem 2.1 of Newey and McFadden (1994, p.2121), their discussion on page 2122 and our Assumptions 1(a)–(c) that

$$\hat{\alpha}_T \xrightarrow{p} \alpha_0, \tag{25}$$

$$\hat{\beta}_T \xrightarrow{p} \beta_0. \tag{26}$$

Thus $\hat{\alpha}_T \in \operatorname{int} A$ and $\hat{\beta}_T \in \operatorname{int} B$ with probability approaching one.

Let

$$B_\epsilon(\hat{\alpha}) = \{\alpha \in A : \|\alpha - \hat{\alpha}_T\| < \epsilon\}$$

where $\epsilon > 0$. Write the marginal likelihood as the sum of two integrals:

$$\begin{aligned}
m_A &= (2\pi/T)^{-\frac{k_A}{2}} |\widehat{W}_{A,T}|^{\frac{1}{2}} \int_{B_\epsilon(\hat{\alpha}_T)} \pi_A(\alpha) e^{-T\hat{q}_{A,T}(\alpha)} d\alpha \\
&\quad + (2\pi/T)^{-\frac{k_A}{2}} |\widehat{W}_{A,T}|^{\frac{1}{2}} \int_{A \backslash B_\epsilon(\hat{\alpha}_T)} \pi_A(\alpha) e^{-T\hat{q}_{A,T}(\alpha)} d\alpha.
\end{aligned} \tag{27}$$

By Taylor's theorem,

$$\hat{q}_{A,T}(\alpha) = \hat{q}_{A,T}(\hat{\alpha}_T) + \nabla \hat{q}_{A,T}(\hat{\alpha}_T)'(\alpha - \hat{\alpha}_T) + \frac{1}{2}(\alpha - \hat{\alpha}_T)' \nabla^2 \hat{q}_{A,T}(\bar{\alpha}_T(\alpha))(\alpha - \hat{\alpha}_T) \tag{28}$$

where $\bar{\alpha}_T(\alpha)$ is a point between α and $\hat{\alpha}_T$. Thus the first integral on the right hand side of (27) can be written as:

$$\begin{aligned}
&\int_{B_\epsilon(\hat{\alpha}_T)} \pi_A(\alpha) e^{-T\hat{q}_{A,T}(\hat{\alpha}_T) - \frac{T}{2}(\alpha - \hat{\alpha}_T)' \nabla^2 \hat{q}_{A,T}(\bar{\alpha}_T(\alpha))(\alpha - \hat{\alpha}_T)} d\alpha \\
&= e^{-T\hat{q}_{A,T}(\hat{\alpha}_T)} \int_{B_\epsilon(\hat{\alpha}_T)} \pi_A(\alpha) e^{-\frac{T}{2}(\alpha - \hat{\alpha}_T)' \nabla^2 \hat{q}_{A,T}(\hat{\alpha}_T)(\alpha - \hat{\alpha}_T)} d\alpha + O_p(\epsilon) \\
&= e^{-T\hat{q}_{A,T}(\hat{\alpha}_T)} \left(\frac{2\pi}{T}\right)^{\frac{p_A}{2}} \left|\nabla^2 \hat{q}_{A,T}(\hat{\alpha}_T)\right|^{-\frac{1}{2}} (1 + o_p(1)) + O_p(\epsilon).
\end{aligned} \tag{29}$$

It follows from Assumption 1(b) that the second integral on the right hand side of (27) can be bounded by

$$\begin{aligned}
\left| \int_{A \backslash B_\epsilon(\hat{\alpha}_T)} \pi_A(\alpha) e^{-T\hat{q}_{A,T}(\alpha)} d\alpha \right| &\leq \int_{A \backslash B_\epsilon(\hat{\alpha}_T)} \pi_A(\alpha) d\alpha e^{-T \inf_{\alpha \in A \backslash B_\epsilon(\hat{\alpha}_T)} \hat{q}_{A,T}(\alpha)} \\
&= O_p\left(e^{-T(\hat{q}_{A,T}(\hat{\alpha}_T) + \eta)}\right)
\end{aligned} \tag{30}$$

for some $\eta > 0$. It follows from (29) and (30) that the marginal likelihood can be approximated by

$$m_A = e^{-T\hat{q}_{A,T}(\hat{\alpha}_T)} \left(\frac{T}{2\pi}\right)^{\frac{k_A - p_A}{2}} |\widehat{W}_{A,T}|^{\frac{1}{2}} \left|\nabla^2 \hat{q}_{A,T}(\hat{\alpha}_T)\right|^{-\frac{1}{2}} (1 + o_p(1)). \qquad (31)$$

Similarly, we obtain

$$m_B = e^{-T\hat{q}_{B,T}(\hat{\alpha}_T)} \left(\frac{T}{2\pi}\right)^{\frac{k_B - p_B}{2}} |\widehat{W}_{B,T}|^{\frac{1}{2}} \left|\nabla^2 \hat{q}_{B,T}(\hat{\beta}_T)\right|^{-\frac{1}{2}} (1 + o_p(1)). \qquad (32)$$

Proof of Theorem 1(a). Without loss of generality, suppose that $q_A(\alpha_0) < q_B(\beta_0)$. It follows from (31) and (32) that

$$
\begin{aligned}
\ln\left(\frac{m_A}{m_B}\right) &= -T(\hat{q}_{A,T}(\hat{\alpha}_T) - \hat{q}_{B,T}(\hat{\beta}_T)) + \frac{k_A - p_A - k_B + p_B}{2} \ln\left(\frac{T}{2\pi}\right) + \ln\left(\frac{\pi_A(\hat{\alpha}_T)}{\pi_B(\hat{\beta}_T)}\right) \\
&\quad + \frac{1}{2}\ln\left(\frac{\left|\widehat{W}_{A,T}\right|}{\left|\widehat{W}_{B,T}\right|}\right) - \frac{1}{2}\ln\left(\frac{\left|\nabla^2\hat{q}_{A,T}(\hat{\alpha}_T)\right|}{\left|\nabla^2\hat{q}_{B,T}(\hat{\beta}_T)\right|}\right) + o_p(1) \\
&= T(q_B(\beta_0) - q_A(\alpha_0)) + o_p(T). \qquad (33)
\end{aligned}
$$

Because $q_B(\beta_0) - q_A(\alpha_0) > 0$, $m_A > m_B$ with probability approaching one.

Proof of Theorem 1(b). Without loss of generality, suppose that $q_A(\alpha_0) = q_B(\beta_0)$ with $k_A - p_A > k_B - p_B$. Then we have

$$
\begin{aligned}
\ln\left(\frac{m_A}{m_B}\right) &= -T(\hat{q}_{A,T}(\hat{\alpha}_T) - \hat{q}_{B,T}(\hat{\beta}_T)) + \frac{k_A - p_A - k_B + p_B}{2} \ln\left(\frac{T}{2\pi}\right) + \ln\left(\frac{\pi_A(\hat{\alpha}_T)}{\pi_B(\hat{\beta}_T)}\right) \\
&\quad + \frac{1}{2}\ln\left(\frac{\left|\widehat{W}_{A,T}\right|}{\left|\widehat{W}_{B,T}\right|}\right) - \frac{1}{2}\ln\left(\frac{\left|\nabla^2\hat{q}_{A,T}(\hat{\alpha}_T)\right|}{\left|\nabla^2\hat{q}_{B,T}(\hat{\beta}_T)\right|}\right) + o_p(1) \\
&= \frac{k_A - p_A - k_B + p_B}{2} \ln(T) + O_p(1). \qquad (34)
\end{aligned}
$$

Because $k_A - p_A - k_B + p_B > 0$, $m_A > m_B$ with probability approaching one.

Proof of Theorem 2(a). The proof of Theorem 2(a) is analogous to that of Theorem 1(a) and thus is omitted.

Proof of Theorem 2(b). When $T^{\frac{1}{2}}(\hat{q}_{A,T}(\hat{\alpha}_T) - \hat{q}_{B,T}(\hat{\beta}_T))$ converges in distribution to a

nondegenerate zero-mean symmetric distribution,

$$
\begin{aligned}
T^{-\frac{1}{2}}\ln\left(\frac{m_A}{m_B}\right) &= -\sqrt{T}(\hat{q}_{A,T}(\hat{\alpha}_T) - \hat{q}_{B,T}(\hat{\beta}_T)) + \frac{k_A - p_A - k_B + p_B}{2}T^{-\frac{1}{2}}\ln\left(\frac{T}{2\pi}\right) + T^{-\frac{1}{2}}\ln\left(\frac{\pi_A(\hat{\alpha}_T)}{\pi_B(\hat{\beta}_T)}\right) \\
&\quad + \frac{1}{2}T^{-\frac{1}{2}}\ln\left(\frac{\left|\widehat{W}_{A,T}\right|}{\left|\widehat{W}_{B,T}\right|}\right) - \frac{1}{2}T^{-\frac{1}{2}}\ln\left(\frac{\left|\nabla^2\hat{q}_{A,T}(\hat{\alpha}_T)\right|}{\left|\nabla^2\hat{q}_{B,T}(\hat{\beta}_T)\right|}\right) + o_p(1) \\
&= -\sqrt{T}(\hat{q}_{A,T}(\hat{\alpha}_T) - \hat{q}_{B,T}(\hat{\beta}_T)) + o_p(1)
\end{aligned}
\tag{35}
$$

also converges in distribution to a nondegenerate zero-mean symmetric distribution. Thus $m_A > m_B$ with probability approaching one half.

Proof of Theorem 3(a). The proof of Theorem 3(a) is similar to that of Theorem 1(a) with T replaced by $T^{\frac{1}{2}}$ and thus is omitted.

Proof of Theorem 3(b). Without loss of generality, suppose that $q_A(\alpha_0) = q_B(\beta_0)$ with $k_A - p_A > k_B - p_B$. Then we have

$$
\begin{aligned}
\ln\left(\frac{\tilde{m}_A}{\tilde{m}_B}\right) &= -\sqrt{T}(\hat{q}_{A,T}(\hat{\alpha}_T) - \hat{q}_{B,T}(\hat{\beta}_T)) + \frac{k_A - p_A - k_B + p_B}{2}\ln\left(\frac{T}{2\pi}\right) + \ln\left(\frac{\pi_A(\hat{\alpha}_T)}{\pi_B(\hat{\beta}_T)}\right) \\
&\quad + \frac{1}{2}\ln\left(\frac{\left|\widehat{W}_{A,T}\right|}{\left|\widehat{W}_{B,T}\right|}\right) - \frac{1}{2}\ln\left(\frac{\left|\nabla^2\hat{q}_{A,T}(\hat{\alpha}_T)\right|}{\left|\nabla^2\hat{q}_{B,T}(\hat{\beta}_T)\right|}\right) + o_p(1) \\
&= \frac{k_A - p_A - k_B + p_B}{2}\ln(T) + O_p(1).
\end{aligned}
\tag{36}
$$

Because $k_A - p_A - k_B + p_B > 0$, we have $m_A > m_B$ with probability approaching one.

We will use the following lemma in the proof of Theorem 4:

Lemma 2. Suppose that Assumption 2 holds. Define a profile estimator of α_s by

$$
\hat{\alpha}_{s,T}(\alpha_w) = \mathrm{argmin}_{\alpha_s \in A_s}(\hat{\gamma}_T - f(\alpha_s, \alpha_w))'\widehat{W}_{A,T}(\hat{\gamma}_T - f(\alpha_s, \alpha_w))
\tag{37}
$$

for each $\alpha_w \in A_w$. Then

$$
\sup_{\alpha_w \in A_w} \|\hat{\alpha}_s(\alpha_w) - \alpha_{s,0}\| = O_p(T^{-\frac{1}{2}}), \tag{38}
$$

$$
\sup_{\alpha_w \in A_w} |\hat{q}_{A,T}(\hat{\alpha}_{s,T}(\alpha_w), \alpha_w) - \hat{q}_{A,T}(\hat{\alpha}_{s,T}(\alpha_{w,0}), \alpha_{w,0})| = O_p(T^{-\frac{1}{2}}), \tag{39}
$$

$$
\sup_{\alpha_w \in A_w} \left\|\mathrm{vech}\left(\nabla^2_{\alpha_s}\hat{q}_{A,T}(\hat{\alpha}_{s,T}(\alpha_w), \alpha_w) - \nabla^2_{\alpha_s}\hat{q}_{A,T}(\hat{\alpha}_{s,T}(\alpha_{w,0}), \alpha_{w,0})\right)\right\| = O_p(T^{-\frac{1}{2}}), \tag{40}
$$

for every $\alpha_{w,0} \in A_w$.

Proof of Lemma 2.

Note that $\hat{\alpha}_{s,T}(\alpha_w)$ satisfies the first order conditions:

$$F_{\alpha_s}(\hat{\alpha}_{s,T}(\alpha_w), \alpha_w)' \widehat{W}_{A,T}(\hat{\gamma}_T - f(\hat{\alpha}_{s,T}(\alpha_w), \alpha_w)) = 0_{p_{A_s} \times 1}, \tag{41}$$

where $F_{\alpha_s}(\alpha) = \partial f(\alpha) / \partial \alpha_s'$. Let $\hat{J}_s$ and $\hat{J}_w$ denote the Jacobian matrices of the left hand side of (41) with respect to α_s and α_w:

$$
\begin{aligned}
\hat{J}_s &= -F_{\alpha_s}(\hat{\alpha}_{s,T}(\alpha_w), \alpha_w)' \widehat{W}_{A,T} F_{\alpha_s}(\hat{\alpha}_{s,T}(\alpha_w), \alpha_w) \\
&\quad + [(\hat{\gamma}_T - f(\hat{\alpha}_{s,T}(\alpha_w), \alpha_w))' \widehat{W}_{A,T} \otimes I] \frac{\partial \text{vec}(F_{\alpha_s}(\hat{\alpha}_{s,T}(\alpha_w), \alpha_w))}{\partial \alpha_s'}, \tag{42} \\
\hat{J}_w &= -T^{-\frac{1}{2}} F_{\alpha_s}(\hat{\alpha}_{s,T}(\alpha_w), \alpha_w)' \widehat{W}_{A,T} F_{\alpha_w}(\hat{\alpha}_{s,T}(\alpha_w), \alpha_w) \\
&\quad + T^{-\frac{1}{2}} [(\hat{\gamma}_T - f(\hat{\alpha}_{s,T}(\alpha_w), \alpha_w))' \widehat{W}_{A,T} \otimes I] \frac{\partial \text{vec}(F_{\alpha_w}(\hat{\alpha}_{s,T}(\alpha_w), \alpha_w))}{\partial \alpha_w'} \\
&= O_p(T^{-\frac{1}{2}}), \tag{43}
\end{aligned}
$$

where $F_{\alpha_w}(\alpha) = \partial f_w / \partial \alpha_w'$ and $O_p(T^{-1/2})$ is uniform in $\alpha_w \in A_w$ because of the compactness of A and twice continuous differentiability of f. By Assumption 2(e), $\hat{J}_s$ is nonsingular with probability approaching one. Thus, it follows from (42), (43) and the implicit function theorem

$$\partial \hat{\alpha}_{s,T}(\alpha_w) / \partial \alpha_w' = -\hat{J}_s^{-1} \hat{J}_w = O_p(T^{-\frac{1}{2}}) \tag{44}$$

where $O_p(T^{-1/2})$ is uniform in $\alpha_w \in A_w$. It follows from the mean value theorem and (44) that

$$\hat{\alpha}_{s,T}(\alpha_w') - \hat{\alpha}_{s,T}(\alpha_w) = O_p(T^{-\frac{1}{2}} \| \alpha_w' - \alpha_w \|). \tag{45}$$

Given the pointwise convergence $\hat{\alpha}_{s,T}(\alpha_w) \xrightarrow{p} \alpha_{s,0}$ for each $\alpha_w \in A_w$ (which is straightforward to show), the compactness of A_w and stochastic equicontinuity (45), we can strengthen the pointwise convergence to uniform convergence (38) by Theorem 1 of Andrews (1992): (39) and (40) follow from (38) and other

assumptions.

Proof of Theorem 4. Let

$$B_\epsilon(\hat\alpha_{s,T}) = \{\alpha_s \in A_s : \|\alpha_s - \hat\alpha_{s,T}\| < \epsilon\}.$$

where $\epsilon > 0$. Then it follows from Assumptions 2(b) and 2(c) that

$$\left| \int_{(A_s \backslash B_\epsilon(\hat\alpha_T)) \times A_w} \pi_A(\alpha) e^{-T\hat q_{A,T}(\alpha)} d\alpha \right| \leq \int_{(A_s \backslash B_\epsilon(\hat\alpha_T)) \times A_w} \pi_A(\alpha) d\alpha e^{-T \inf_{\alpha \in (A_s \backslash B_\epsilon(\hat\alpha_T)) \times A_w} \hat q_{A,T}(\alpha)}$$

$$= O\left(e^{-T(\hat q_{A,T}(\hat\alpha_T) + \eta)}\right), \tag{46}$$

for some $\eta > 0$. Let $\alpha_{w,0} \in A_w$. Using arguments similar to the one used to obtain (30), we can write

$$\int_{B_\epsilon(\hat\alpha_{s,T})} \pi_A(\alpha) e^{-T\hat q_{A,T}(\alpha)} d\alpha$$

$$= \int_{B_\epsilon(\hat\alpha_{s,T})} \pi_A(\alpha) e^{-T\hat q_{A,T}(\hat\alpha_{s,T}(\alpha_w), \alpha_w)} e^{-\frac{T}{2}(\alpha_s - \hat\alpha_{s,T}(\alpha_w))' \nabla^2_{\alpha_s} \hat q_{A,T}(\hat\alpha_{s,T}(\alpha_w), \alpha_w)(\alpha_s - \hat\alpha_{s,T}(\alpha_w))} d\alpha + O_p(\epsilon),$$

$$= \int_{B_\epsilon(\hat\alpha_{s,T})} \pi_A(\alpha) e^{-T\hat q_{A,T}(\hat\alpha_{s,T}(\alpha_{w,0}), \alpha_{w,0})}$$

$$\times e^{-\frac{T}{2}(\alpha_s - \hat\alpha_{s,T}(\alpha_{w,0}))' \nabla^2_{\alpha_s} \hat q_{A,T}(\hat\alpha_{s,T}(\alpha_{w,0}), \alpha_{w,0})(\alpha_s - \hat\alpha_{s,T}(\alpha_{w,0}))} d\alpha(1 + o_p(1)) + O_p(\epsilon)$$

$$= e^{-T\hat q_{A,T}(\hat\alpha_{s,T}(\alpha_{w,0}), \alpha_{w,0})} \left(\frac{2\pi}{T}\right)^{\frac{p_{A_s}}{2}} \left\|\nabla^2_{\alpha_s} \hat q_{A,T}(\hat\alpha_{s,T}(\alpha_{w,0}), \alpha_{w,0})\right\|^{-\frac{1}{2}} \int_{A_w} \pi_A(\alpha_{s,0}, \alpha_{w,0})(1 + o_p(1)) + O_p(\epsilon),$$

$$\tag{47}$$

where the second equality follows from (38)–(40). Combining (46) and (47) we obtain

$$m_A = e^{-T\hat q_{A,T}(\hat\alpha_{s,T}(\alpha_{w,0}), \alpha_{w,0})} \left(\frac{T}{2\pi}\right)^{\frac{k_A - p_{A_s}}{2}}$$

$$\times \left|\widehat W_{A,T}\right|^{\frac{1}{2}} \left|\nabla^2_{\alpha_s} \hat q_{A,T}(\hat\alpha_{s,T}(\alpha_{w,0}), \alpha_{w,0}\right|^{-\frac{1}{2}} \int_{A_w} \pi_A(\alpha_{s,0}, \alpha_w)(1 + o_p(1)) \tag{48}$$

When there is no strongly identified parameter ($p_{A_s} = 0$), we have

$$m_A = e^{-\frac{T}{2}\hat\gamma'_{A,T} \widehat W_{A,T} \hat\gamma_{A,T}}(1 + o_p(1)) \tag{49}$$

and $\hat\gamma'_T \widehat W_T \hat\gamma_T$ plays the same role as $\hat q_{A,T}(\hat\alpha_{s,T}(\alpha_{w,0}), \alpha_{w,0})$ in (48). Using arguments similar to those used in the proof of Theorems 1, 2 and 3, Theorem 4(a),(b) and (c) follow from (48) and (49).

27

Proof of Theorem 5(a).

We can write

$$
\begin{aligned}
\int_A \pi_A(\alpha) \exp\left(-T\hat{q}_{A,T}(\alpha)\right) d\alpha &= \int_{A_0} \pi_A(\alpha) \exp\left(-T\hat{q}_{A,T}(\alpha)\right) d\alpha \\
&\quad + \int_{(A_0^c)^{-\varepsilon}} \pi_A(\alpha) \exp\left(-T\hat{q}_{A,T}(\alpha)\right) d\alpha \\
&\quad + \int_{A\backslash(A_0\cup(A_0^c)^{-\varepsilon})} \pi_A(\alpha) \exp\left(-T\hat{q}_{A,T}(\alpha)\right) d\alpha \\
&= I_1 + I_2 + I_3, \quad \text{say.} \tag{50}
\end{aligned}
$$

It follows from Assumptions 3(a)(c) that

$$
\begin{aligned}
I_1 &= \int_{A_0} \pi_A(\alpha) \exp\left(-Tq_A(\alpha)\right) d\alpha + o_p\left(\int_{A_0} \pi_A(\alpha) \exp\left(-Tq_A(\alpha)\right) d\alpha\right) \\
&= \int_{A_0} \pi_A(\alpha) d\alpha \exp\left(-Tq_A(\alpha_0)\right) d\alpha + o_p\left(\exp\left(-Tq_A(\alpha_0)\right)\right), \tag{51}
\end{aligned}
$$

for any $\alpha_0 \in A_0$. It follows from Assumptions 3(a)(b)(c) that

$$
I_2 = o_p\left(\exp\left(-Tq_A(\alpha_0)\right)\right). \tag{52}
$$

By letting $\varepsilon \to 0$, the term I_3 can be made arbitrarily small. Combining (50)–(52), we can approximate the quasi marginal likelihood for model A by

$$
m_A = \left(\frac{T}{2\pi}\right)^{\frac{k_A}{2}} \left|\widehat{W}_{A,T}\right|^{\frac{1}{2}} \int_{A_0} \pi_A(\alpha) d\alpha \exp\left(-Tq_A(\alpha_0)\right) + o_p\left(T^{\frac{k_A}{2}} \exp\left(-Tq_A(\alpha_0)\right)\right). \tag{53}
$$

Similarly, the quasi marginal likelihood for model B can be approximated by

$$
m_B = \left(\frac{T}{2\pi}\right)^{\frac{k_B}{2}} \left|\widehat{W}_{B,T}\right|^{\frac{1}{2}} = \int_{B_0} \pi_B(\beta) d\beta \exp\left(-Tq_B(\beta_0)\right) d\beta + o_p\left(T^{\frac{k_B}{2}} \exp\left(-Tq_B(\beta_0)\right)\right), \tag{54}
$$

for any $\beta_0 \in B_0$. Theorem 5(a) follows from (53) and (54).

To prove Theorem 5(b) and 5(c), we need the following lemma:

Lemma 3. Suppose that Assumption 3 holds. Define a profile estimator of α_s by

$$
\hat{\alpha}_{s,T}(\alpha_p) = \text{argmin}_{\alpha_s \in A_s} (\hat{\gamma}_{A,T} - f(\alpha_s, \alpha_p))' \widehat{W}_{A,T} (\hat{\gamma}_{A,T} - f(\alpha_s, \alpha_p)) \tag{55}
$$

for each $\alpha_p \in A_p$. Then

$$\sup_{\alpha_p \in A_{p,0}} \|\hat{\alpha}_s(\alpha_p) - \alpha_{s,0}\| = o_p(1), \quad (56)$$

$$\sup_{\alpha_p \in A_{p,0}} |\hat{q}_{A,T}(\hat{\alpha}_{s,T}(\alpha_p), \alpha_p) - \hat{q}_{A,T}(\hat{\alpha}_{s,T}(\alpha_{p,0}), \alpha_{p,0})| = o_p(1), \quad (57)$$

$$\sup_{\alpha_p \in A_{p,0}} \left\| \text{vech}\left(\nabla^2_{\alpha_s} \hat{q}_{A,T}(\hat{\alpha}_{s,T}(\alpha_p), \alpha_p) - \nabla^2_{\alpha_s} \hat{q}_{A,T}(\hat{\alpha}_{s,T}(\alpha_{p,0}), \alpha_{p,0})\right)\right| = o_p(1), \quad (58)$$

for any $\alpha_{p,0} \in A_{p,0}$.

Proof of Lemma 3.

Note that $\hat{\alpha}_{s,T}(\alpha_p)$ satisfies the first order conditions:

$$F_{\alpha_s}(\hat{\alpha}_{s,T}(\alpha_p), \alpha_p)' \widehat{W}_{A,T}(\hat{\gamma}_{A,T} - f(\hat{\alpha}_{s,T}(\alpha_p), \alpha_p)) = 0_{p_{A_s} \times 1}, \quad (59)$$

for every $a_p \in A_{p,0}$. Let $\hat{J}_s$ and $\hat{J}_p$ denote the Jacobian matrices of the left hand side of (59) with respect to α_s and α_p are

$$\hat{J}_s = -F_{\alpha_s}(\hat{\alpha}_{s,T}(\alpha_p), \alpha_p)' \widehat{W}_{A,T} F_{\alpha_s}(\hat{\alpha}_{s,T}(\alpha_p), \alpha_p)$$
$$+ [(\hat{\gamma}_{A,T} - f(\hat{\alpha}_{s,T}(\alpha_p), \alpha_p))' \widehat{W}_{A,T} \otimes I] \frac{\partial \text{vec}(F_{\alpha_s}(\hat{\alpha}_{s,T}(\alpha_p), \alpha_p))}{\partial \alpha'_s}, \quad (60)$$

$$\hat{J}_p = -F_{\alpha_s}(\hat{\alpha}_{s,T}(\alpha_p), \alpha_p)' \widehat{W}_{A,T} F_{\alpha_p}(\hat{\alpha}_{s,T}(\alpha_p), \alpha_p)$$
$$+ [(\hat{\gamma}_{A,T} - f(\hat{\alpha}_{s,T}(\alpha_p), \alpha_p))' \widehat{W}_{A,T} \otimes I] \frac{\partial \text{vec}(F_{\alpha_p}(\hat{\alpha}_{s,T}(\alpha_p), \alpha_p))}{\partial \alpha'_p}, \quad (61)$$

where $F_{\alpha_p}(\alpha) = \partial f / \partial \alpha'_p$. because of the compactness of A and twice continuous differentiability of f. By Assumption 3(e), $\hat{J}_s$ is nonsingular with probability approaching one. Thus, it follows from (60), (61) and the implicit function theorem

$$\partial \hat{\alpha}_{s,T}(\alpha_p) / \partial \alpha'_p = -\hat{J}_s^{-1} \hat{J}_p = O_p(1). \quad (62)$$

It follows from the mean value theorem and (62) that

$$\hat{\alpha}_{s,T}(\alpha'_p) - \hat{\alpha}_{s,T}(\alpha_p) = O_p(\|\alpha'_p - \alpha_p\|), \quad (63)$$

where $\alpha_p, \alpha'_p \in A_{p,0}$. Given the pointwise convergence $\hat{\alpha}_{s,T}(\alpha_p) \overset{p}{\to} \alpha_{s,0}$ for each $\alpha_p \in A_{p,0}$ (which is straightforward to show), the compactness of A_p and stochastic equicontinuity (63), we can strengthen the pointwise convergence to

uniform convergence (56) by Theorem 1 of Andrews (1992): (57) and (58) follow from (56) and other assumptions.

Proof of Theorem 5(b).

Define

$$B_\epsilon(\hat{\alpha}_{s,T}) = \{\alpha_s \in A_s : \|\alpha_s - \hat{\alpha}_{s,T}\|\}.$$

The quasi-marginal likelihood of model A can be written as

$$
\begin{aligned}
&\int_A \pi_A(\alpha) \exp\left(-T\hat{q}_{A,T}(\alpha)\right) d\alpha \\
&= \int_{B_\epsilon(\hat{\alpha}_{s,T}) \times A_{p,0}} \pi_A(\alpha) \exp\left(-T\hat{q}_{A,T}(\alpha)\right) d\alpha \\
&\quad + \int_{((B_\epsilon(\hat{\alpha}_{s,T}) \times A_{p,0})^c)^{-\varepsilon}} \pi_A(\alpha) \exp\left(-T\hat{q}_{A,T}(\alpha)\right) d\alpha \\
&\quad + \int_{A \setminus ((B_\epsilon(\hat{\alpha}_{s,T}) \times A_0) \cup (B_\epsilon(\hat{\alpha}_{s,T}) \times A_0)^c))^{-\varepsilon})} \pi_A(\alpha) \exp\left(-T\hat{q}_{A,T}(\alpha)\right) d\alpha \\
&= I_1 + I_2 + I_3, \quad \text{say.}
\end{aligned}
\tag{64}
$$

Given $\alpha_{p,0} \in A_{p,0}$, we can write

$$
\begin{aligned}
I_1 &= \int_{B_\epsilon(\hat{\alpha}_{s,T}) \times A_{p,0}} \pi_A(\alpha) e^{-T\hat{q}_{A,T}(\alpha)} d\alpha \\
&= \int_{B_\epsilon(\hat{\alpha}_{s,T})} \pi_A(\alpha) e^{-T\hat{q}_{A,T}(\hat{\alpha}_{s,T}(\alpha_p),\alpha_p)} e^{-\frac{T}{2}(\alpha_s - \hat{\alpha}_{s,T}(\alpha_p))' \nabla^2_{\alpha_s} \hat{q}_{A,T}(\hat{\alpha}_{s,T}(\alpha_p),\alpha_p)(\alpha_s - \hat{\alpha}_{s,T}(\alpha_p))} d\alpha + O_p(\epsilon), \\
&= \int_{B_\epsilon(\hat{\alpha}_{s,T})} \pi_A(\alpha) e^{-T\hat{q}_{A,T}(\hat{\alpha}_{s,T},\alpha_{p,0})} e^{-\frac{T}{2}(\alpha_s - \hat{\alpha}_{s,T}(\alpha_{p,0}))' \nabla^2_{\alpha_s} \hat{q}_{A,T}(\hat{\alpha}_{s,T}(\alpha_{p,0}),\alpha_{p,0})(\alpha_s - \hat{\alpha}_{s,T}(\alpha_{p,0}))} d\alpha + O_p(\epsilon), \\
&= e^{-T\hat{q}_{A,T}(\hat{\alpha}_{s,T}(\alpha_{p,0}),\alpha_{p,0})} \left(\frac{2\pi}{T}\right)^{\frac{p_{A_s}}{2}} \left\|\nabla^2_{\alpha_s} \hat{q}_{A,T}(\hat{\alpha}_{s,T}(\alpha_{p,0}),\alpha_{p,0})\right\|^{-\frac{1}{2}} \int_{A_{p,0}} \pi_A(\alpha_{s,0},\alpha_p) d\alpha_p (1 + o_p(1)) \\
&\quad + O_p(\epsilon),
\end{aligned}
\tag{65}
$$

where the second last equality follows from Lemma 3. Thus, as in Theorem 5(a), the quasi-marginal likelihood can be approximated by

$$
\begin{aligned}
m_A &= e^{-T\hat{q}_{A,T}(\hat{\alpha}_{s,T}(\alpha_{p,0}),\alpha_{p,0})} \left(\frac{T}{2\pi}\right)^{\frac{k_A - p_{A_s}}{2}} \left|\widehat{W}_{A,T}\right|^{\frac{1}{2}} \\
&\quad \times \left|\nabla^2_{\alpha_s} \hat{q}_{A,T}(\hat{\alpha}_{s,T}(\alpha_{p,0}),\alpha_{p,0})\right|^{-\frac{1}{2}} \int_{A_{p,0}} \pi_A(\alpha_{s,0},\alpha_p) d\alpha_p (1 + o_p(1))
\end{aligned}
\tag{66}
$$

The rest of the proof is analogous to those of the previous theorems.

References

Altig, David E., Lawrence J. Christiano, Martin Eichenbaum and Jesper Lindé (2011), "Firm-Specific Capital, Nominal Rigidities and the Business Cycle," *Review of Economic Dynamics*, 14 (2), 225–247.

An, Sungbae, and Frank Schorfheide (2007), "Bayesian Analysis of DSGE Models," *Econometric Reviews,* 26, 113–172.

Andrews, Donald W.K., (1991), "Generic Uniform Convergence," *Econometric Theory,* 8, 241–257.

Andrews, Donald W.K., (1999), "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation," *Econometrica,* 67, 543–564.

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

Canova, Fabio, and Luca Sala (2009), "Back to Square One: Identification Issues in DSGE Models," *Journal of Monetary Economics*, 56, 431–449.

Chernozhukov, Victor, and Han Hong (2003), "An MCMC Approach to Classical Estimation," *Journal of Econometrics,* 293–346.

Chernozhukov, Victor, Han Hong and Elie Tamer (2007), "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284.

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

Christiano, Lawrence J., (1991), "Modeling the Liquidity Effect of a Money Shock," *Federal Reserve Bank of Minneapolis Quarterly Review*, 15 (1), 3–34.

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

Christiano, Lawrence J., Martin S. Eichenbaum and Mathias Trabandt (2013), "Unemployment and Business Cycles," unpublished manuscript, Northwestern University and Board of Governors of the Federal Reserve System.

Christiano, Lawrence J., Mathias Trabandt and Karl Walentin (2011), "DSGE Models for Monetary Policy Analysis," in Benjamin M. Friedman and Michael Woodford, editors: *Handbook of Monetary Economics,* Volume 3A, The Netherlands: North-Holland.

Corradi, Valentina, and Norman R. Swanson (2007), "Evaluation of Dynamic Stochastic General Equilibrium Models Based on Distributional Comparison of Simulated and Historic Data," *Journal of Econometrics*, 136 (2), 699–723.

Del Negro, Marco, and Frank Schorfheide (2004), "Priors from General Equilibrium Models for VARs," *International Economic Review*, 45(2), 643–73.

Del Negro, Marco, and Frank Schorfheide (2009), "Monetary Policy Analysis with Potentially Misspecified Models," *American Economic Review*, 99 (4), 1415–1450.

Dridi, Ramdan, Alain Guay and Eric Renault (2007), "Indirect Inference and Calibration of Dynamic Stochastic General Equilibrium Models," *Journal of Econometrics*, 136 (2), 397–430.

Geweke, John, (1998), "Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication," Staff Report 249, Federal Reserve Bank of Minneapolis.

Guerron-Quintana, Pablo, Atsushi Inoue and Lutz Kilian (2013), "Frequentist Inference in Weakly Identified DSGE Models," *Quantitative Economics*, forthcoming.

Inoue, Atsushi, and Lutz Kilian (2006), "On the Selection of Forecasting Models," *Journal of Econometrics,* 130, 273–306.

Jòrda, Òscar, and Sharon Kozicki (2011), "Estimation and Inference by the Method of Projection Minimum Distance," *International Economic Review,* 52, 461–487.

Kim, Jae-Young (2000), "Limited Information Likelihood and Bayesian Analysis," *Journal of Econometrics*, 107 , 175–193.

Kormilitsina, Anna, and Denis Nekipelov (2013), "Consistent Variance of the Laplace Type Estimators: Application to DSGE Models," unpublished manuscript, Southern Methodist University and University of California, Berkeley.

Liao, Yuan and Wenxin Jiang (2010), "Bayesian Analysis in Moment Inequality Models," *Annals of Statistics,* 38, 275–316.

Moon, Hyungsik Roger, and Frank Schorfheide (2012), "Bayesian and Frequentist Inference in Partially Identified Models," *Econometrica*, 80, 755–782.

Nason, James M., and Timothy Cogley (1994), "Testing the Implications of Long-Run Neutrality for Monetary Business Cycle Models," *Journal of Applied Econometrics*, 9 (S1), S37–S70.

Newey, Whitney K., and Daniel L. McFadden (1994), "Large Sample Estimation and Hypothesis Testing," in Robert F. Engle and Daniel L. McFadden eds., *Handbook of Econometrics,* Volume IV, 2111–2245.

Phillips, Peter C.B. (1996), "Econometric Model Determination," *Econometrica*, 64, 763–812.

Schorfheide, Frank, (2000), "Loss Function-Based Evaluation of DSGE Models," *Journal of Applied Econometrics*, 15 (6), 645–670.

Sin, Chor-Yiu, and Halbert White (1996), "Information Criteria for Selecting Possibly Misspecified Parametric Models," *Journal of Econometrics*, 71, 207–225.

Smets, Frank, and Rafael Wouters (2007), "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97, 586–606.

Stock, James H., and Jonathan H. Wright (2000), "GMM with Weak Identification," *Econometrica,* 68, 1055–1096.

Vuong, Quang H., (1989), "Likelihood Ratio Tests for Model Selection and Non-Nested Hypothesis," *Econometrica,* 57, 307–333.

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

Zellner, Arnold (1998), "Past and Recent Results on Maximal Data Information Priors," *Journal of Statistical Research,* 32 (1), 1–22.

Table 1: Simulation Design

Identification	Design	True Model		Model A		Model B	
		σ	κ	σ	κ	σ	κ
Strong	1	1	0.5	estimated	estimated	estimated	fixed at 1
	2	1	0.5	estimated	fixed at 0.5	estimated	estimated
Weak	3	1	0.5	estimated	estimated	estimated	fixed at 1
	4	1	0.5	estimated	fixed at 0.5	estimated	estimated
Partial	5	1	$\alpha = 0.5$ $\zeta = 1$	fixed at 1	estimated (α, ζ)	fixed at 0.5	estimated (α, ζ)
	6	1	$\alpha = 0.5$ $\zeta = 1$	fixed at 1	estimated (α, ζ)	estimated	estimated (α, ζ)

Notes. In the cases of strong and partial identification (design 1,2,5 and 6),

$$f(\sigma, \kappa) \;=\; [1 + \sigma^2,\; \kappa + \sigma^2\kappa,\; -\sigma,\; 1 + \kappa^2 + \sigma^2\kappa^2,\; -\sigma\kappa]',$$

and the corresponding elements of the covariance matrix are used. In the cases of weak identification (designs 3 and 4),

$$f(\sigma, \kappa) \;=\; [\kappa + \sigma^2\kappa,\; 1 + \kappa^2 + \sigma^2\kappa^2,\; -\sigma\kappa]'$$

and the corresponding elements of the covariance matrix are used instead. In the cases of partial identification (designs 5 and 6), $\kappa = (1 - \alpha)(1 - 0.99\alpha)\zeta/\alpha$. Model A is correctly specified while Model B is misspecified in designs 1, 3 and 5. Models A and B are both correctly specified and Model A is more parsimonious than Model B in designs 2, 4 and 6.

Table 2: Frequencies of Selecting Model A: Strong Identification

Quasi-Marginal Likelihood

Design	T	Laplace Approximation		Modified Harmonic Mean	Chib-Jeliazkov	
		Diagonal	Optimal	Optimal	Analytical Optimal	Estimated Optimal
1	50	0.997	0.999	0.999	0.999	0.999
	100	1.000	1.000	1.000	1.000	1.000
	200	1.000	1.000	1.000	1.000	1.000
2	50	0.879	0.840	0.839	0.840	0.830
	100	0.900	0.899	0.899	0.900	0.898
	200	0.940	0.944	0.943	0.943	0.941

Modified Quasi-Marginal Likelihood

Design	T	Laplace Approximation		Modified Harmonic Mean	Chib-Jeliazkov	
		Diagonal	Optimal	Optimal	Analytical Optimal	Estimated Optimal
1	50	0.544	0.706	0.736	0.738	0.770
	100	0.586	0.752	0.770	0.770	0.793
	200	0.718	0.894	0.902	0.903	0.914
2	50	1.000	0.998	0.998	0.998	0.998
	100	1.000	1.000	1.000	1.000	1.000
	200	1.000	1.000	1.000	1.000	1.000

Notes: See Table 1 for the experimental designs. "Optimal" refers to cases in which the weighting matrix is set to the inverse of the bootstrap covariance matrix of impulse responses. "Diagonal" refers to cases in which the weighting matrix is diagonal and their diagonal elements are the reciprocals of the bootstrap variances of impulse responses. "Analytical" refers to cases in which the actual covariance matrix of the proposal density is used in the construction of the Chib-Jeliazkov estimator. "Simulated" refers to cases in which the covariance matrix estimated from the posterior draws is used in the construction of the Chib-Jeliazkov estimator.

Table 3: Frequencies of Selecting Model A: Weak Identification

		Quasi-Marginal Likelihood					Modified Quasi-Marginal Likelihood				
		Laplace Approximation		Modified Harmonic Mean	Chib-Jeliazkov		Laplace Approximation		Modified Harmonic Mean	Chib-Jeliazkov	
					Analytical	Estimated				Analytical	Estimated
Design	T	Diagonal	Optimal	Optimal	Optimal	Optimal	Diagonal	Optimal	Optimal	Optimal	Optimal
3	50	0.940	0.960	0.972	0.973	0.982	0.730	0.728	0.883	0.887	0.952
	100	0.982	0.990	0.992	0.992	0.994	0.621	0.718	0.828	0.841	0.911
	200	0.999	1.000	1.000	1.000	1.000	0.537	0.725	0.864	0.867	0.922
4	50	0.815	0.758	0.690	0.675	0.498	0.967	0.982	0.936	0.939	0.789
	100	0.876	0.823	0.805	0.800	0.746	0.992	1.000	0.999	0.997	0.981
	200	0.956	0.893	0.883	0.883	0.871	0.999	1.000	1.000	1.000	1.000

Notes: See Table 1 for the experimental designs. "Optimal" refers to cases in which the weighting matrix is set to the inverse of the bootstrap covariance matrix of impulse responses. "Diagonal" refers to cases in which the weighting matrix is diagonal and their diagonal elements are the reciprocals of the bootstrap variances of impulse responses. "Analytical" refers to cases in which the actual covariance matrix of the proposal density is used in the construction of the Chib-Jeliazkov estimator. "Simulated" refers to cases in which the covariance matrix estimated from the posterior draws is used in the construction of the Chib-Jeliazkov estimator.

Table 4: Frequencies of Selecting Model A: Partial Identification

| Design | T | Quasi-Marginal Likelihood | | | | | Modified Quasi-Marginal Likelihood | | | | |
| | | Laplace Approximation | | Modified Harmonic Mean | Chib-Jeliazkov | | Laplace Approximation | | Modified Harmonic Mean | Chib-Jeliazkov | |
		Diagonal	Optimal	Optimal	Analytical Optimal	Estimated Optimal	Diagonal	Optimal	Optimal	Analytical Optimal	Estimated Optimal
5	50	NA	NA	0.690	0.663	0.642	NA	NA	0.370	0.360	0.356
	100	NA	NA	0.862	0.817	0.819	NA	NA	0.441	0.412	0.396
	200	NA	NA	0.971	0.966	0.966	NA	NA	0.502	0.476	0.470
6	50	NA	NA	0.926	0.714	0.910	NA	NA	0.990	0.762	0.984
	100	NA	NA	0.965	0.745	0.943	NA	NA	0.997	0.783	0.990
	200	NA	NA	0.985	0.776	0.975	NA	NA	0.999	0.799	0.996

Notes: See Table 1 for the experimental designs. "Optimal" refers to cases in which the weighting matrix is set to the inverse of the bootstrap covariance matrix of impulse responses. "Diagonal" refers to cases in which the weighting matrix is diagonal and their diagonal elements are the reciprocals of the bootstrap variances of impulse responses. "Analytical" refers to cases in which the actual covariance matrix of the proposal density is used in the construction of the Chib-Jeliazkov estimator. "Simulated" refers to cases in which the covariance matrix estimated from the posterior draws is used in the construction of the Chib-Jeliazkov estimator.

Table 5: Prior and Posteriors of Parameters of Baseline CEE Model

| Parameter | | Prior | | | Quasi-posterior | |
		Dist.	Mean	Std	Mean	[5%, 95%]
Price-setting rule						
Price stickiness	ξ_p	Beta	0.50	0.15	0.66	[0.60, 0.72]
Price indexation	ι_p	Beta	0.50	0.15	0.49	[0.32, 0.72]
Wage stickiness	ξ_w	Beta	0.50	0.15	0.85	[0.83, 0.87]
Wage indexation	ι_w	Beta	0.50	0.15	0.30	[0.11, 0.46]
Monetary policy rule						
Interest smoothing	ρ_R	Beta	0.70	0.15	0.89	[0.88, 0.91]
Inflation coefficient	r_π	Gamma	1.70	0.15	1.51	[1.37, 1.65]
GDP coefficient	r_y	Gamma	0.10	0.05	0.15	[0.10, 0.19]
Preference and technology						
Consumption habit	b	Beta	0.50	0.15	0.75	[0.73, 0.78]
Inverse labor supply elast	ϕ	Gamma	1.00	0.50	0.14	[0.04, 0.25]
Capital share	α	Beta	0.25	0.05	0.25	[0.22, 0.28]
Cap util adjustment cost	σ_a	Gamma	0.50	0.30	0.32	[0.23, 0.46]
Investment adjustment cost	S''	Gamma	8.00	2.00	10.4	[8.30, 12.9]
Shocks						
Autocorr invest tech	ρ_Ψ	Beta	0.75	0.15	0.55	[0.42, 0.61]
Std dev neutral tech shock	σ_Z	InvGamma	0.20	0.10	0.23	[0.21, 0.26]
Std dev invest tech shock	σ_Ψ	InvGamma	0.20	0.10	0.17	[0.15, 0.20]
Std dev monetary shock	σ_R	InvGamma	0.40	0.20	0.48	[0.43, 0.54]

Note: Quasi-posterior distribution is evaluated using the random walk Metropolis-Hastings algorithm.

Table 6: Empirical Importance of the Nominal and Real Frictions

| | Base | Nominal frictions | | | | Real frictions | | |
		ξ_p=0.1	ξ_w=0.1	ι_p=0.01	ι_w=0.01	S''=2	b=0.1	σ_a=0.1
Quasi-marginal likelihood								
Laplace	**370**	**341**	**146**	**369**	**373**	**327**	**279**	**366**
MHM	**366**	**340**	**143**	**368**	**371**	**326**	**276**	**364**
Quasi-posterior mean								
ξ_p	0.66	**0.10**	0.95	0.68	0.67	0.74	0.68	0.66
ι_p	0.49	0.53	0.69	**0.01**	0.51	0.48	0.52	0.52
ξ_w	0.85	0.88	**0.10**	0.85	0.87	0.80	0.86	0.85
ι_w	0.30	0.32	0.53	0.34	**0.01**	0.43	0.37	0.29
S''	10.4	10.3	2.74	9.37	9.23	**2.00**	8.07	9.81
b	0.75	0.74	0.53	0.76	0.75	0.69	**0.10**	0.75
σ_a	0.32	0.44	0.62	0.35	0.32	0.39	0.26	**0.10**
SW	-923	-975	-973	-918	-927	-1084	-959	-949

Note: QMLs based on Laplace approximation (Laplace approx.) and modified harmonic mean (MHM) estimator.

Figure 1: Impulse Responses to a Monetary Policy Shock

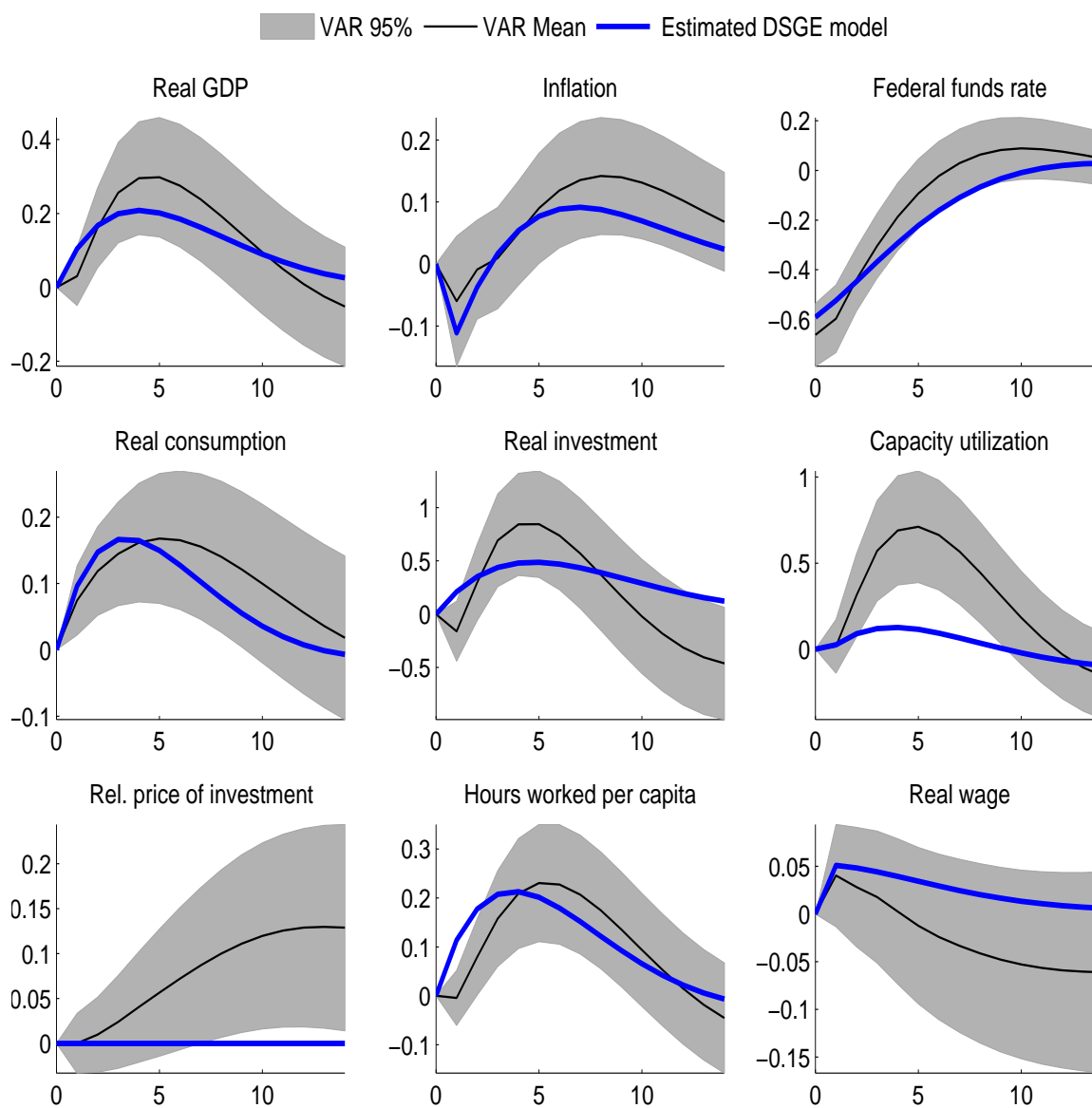VAR 95%   —— VAR Mean   —— Estimated DSGE model

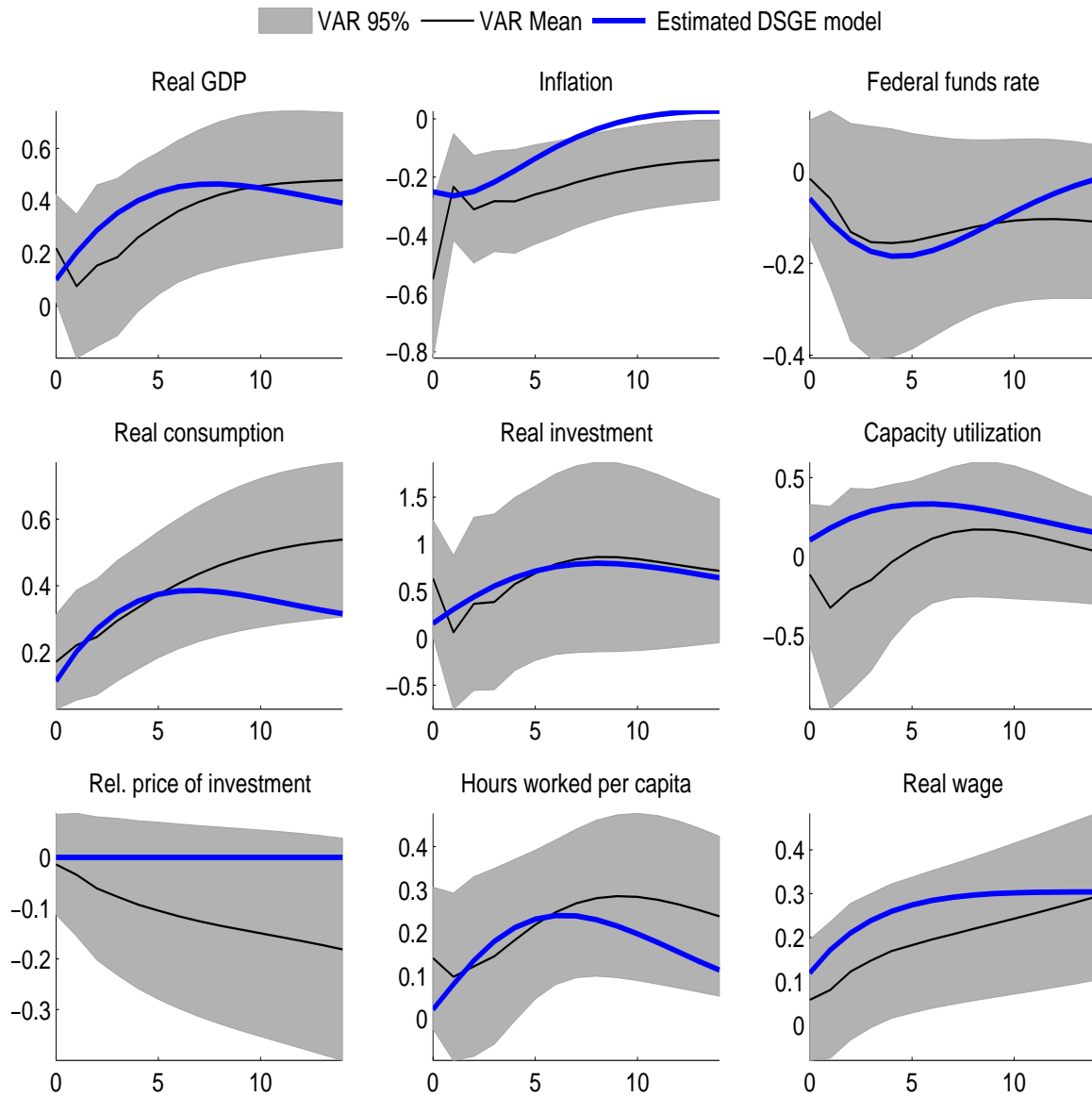Figure 2: Impulse Responses to a Neutral Technology Shock

Figure 3: Impulse Responses to an Investment-specific Technology Shock