

Quasi-Bayesian Model Selection*

Atsushi Inoue[†] Mototsugu Shintani[‡]

Vanderbilt University The University of Tokyo

February 2018

Abstract

In this paper we establish the consistency of the model selection criterion based on the quasi-marginal likelihood (QML) obtained from Laplace-type estimators. 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 New Keynesian macroeconomic models using US data.

*We thank Frank Schorfheide and two anonymous referees for constructive comments and suggestions. We thank Matias Cattaneo, Larry Christiano, Yasufumi Gemma, Kengo Kato, Lutz Kilian, Takushi Kurozumi, Jae-Young Kim and Vadim Marmer for helpful discussions and Mathias Trabandt for providing the data and code. We also thank the seminar and conference participants for helpful comments at the Bank of Canada, Gakushuin University, Hitotsubashi University, Kyoto University, Texas A&M University, University of Tokyo, University of Michigan, Vanderbilt University, 2014 Asian Meeting of the Econometric Society and the FRB Philadelphia/NBER Workshop on Methods and Applications for DSGE Models. Shintani gratefully acknowledges the financial support of Grant-in-aid for Scientific Research.

[†]Department of Economics, Vanderbilt University, 2301 Vanderbilt Place, Nashville, TN 37235. Email: atsushi.inoue@vanderbilt.edu.

[‡]RCAST, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan. Email: moto.shintani@gmail.com.

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 (IRF) 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 The quasi-Bayesian approach does not require the complete specification of likelihood functions and may be robust to potential misspecification. IRF matching can also be used even when shocks are fewer than observed variables (see Fernández-Villaverde, Rubio-Ramírez and Schorfheide, 2016, p.686). Other applications of LTEs to estimate macroeconomic models include Christiano, Eichenbaum and Trabandt (2016), Kormilitsina and Nekipelov (2016), Gemma, Kurozumi and Shintani (2017) and Miyamoto and Nguyen (2017).

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. Likewise, 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 (2016, Table 3) 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

¹The term “quasi-Bayesian” also refers to the procedure that involves data-dependent prior or multiple priors in the Bayesian literature.

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 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 (2012).

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 (QMLs) even if all the competing DSGE models are misspecified.²

The econometric literature on comparing DSGE models include Corradi and Swanson (2007), Dridi, Guay and Renault (2007) and Hnatkovska, Marmer and Tang (2012) who propose hypothesis testing procedures to evaluate the relative performance of possibly misspecified DSGE models. We propose a model selection procedure as in Fernández-Villaverde and Rubio-Ramírez (2004), Hong and Preston (2012) and Kim (2014). In the likelihood framework, Fernández-Villaverde and Rubio-Ramírez (2004)

²As 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.

and Hong and Preston (2012) consider asymptotic properties of the Bayes factor and posterior odds ratio for model comparison, respectively. In the LTE framework, Kim (2014) shows the consistency of the QML criterion for nested model comparison, to which Hong and Preston (2012, p.365) also allude. In a recent paper, Shin (2014) proposes a Bayesian generalized method of moments (GMM) and develops a novel method for computing the marginal likelihood.

We make general contributions in three ways. First, we show that the naive QML model selection criterion may be inconsistent when models are not nested. This is why the existing literature, such as Kim (2014), focuses on the nested case. Second, we develop a new modified QML model selection criterion which remains consistent when nonnested models are considered. Third, we consider cases in which some parameters are either weakly or partially identified. The weakly and partially identified cases are relevant for the estimation of DSGE models but have not been considered in the aforementioned literature.

The outline of this paper is as follows. We begin our analysis by providing a simple illustrative example of model selection in Section 2. Asymptotic justifications for the QML model selection criterion are established in Section 3. We discuss various aspects of implementing IRF matching in Section 4 and provide guidance for the practical implementation of our procedure in Section 5. A small set of Monte Carlo experiments is provided in Section 6. Empirical applications of our procedure to evaluate New Keynesian macroeconomic models using US data are provided in Section 7. The concluding remarks are made in Section 8. The proofs of the theoretical results and additional Monte Carlo results are relegated to the online appendix (Inoue and Shintani, 2017). Throughout the paper, all the asymptotic statements are made for the case in which the sample size tends to infinity or $T \rightarrow \infty$.

2 An Illustrative Example

There are several issues that may arise in model selection. For example, one may compare a correctly specified model and a misspecified model. Or one may compare two correctly specified models where one is more parsimonious than the other. To motivate our proposed QML, we illustrate these issues in a simple Monte Carlo setup and show

that comparing values of estimation criterion functions alone does not necessarily select the preferred model.

Consider a simplified version of the model in Canova and Sala (2009):

$$y_t = E_t(y_{t+1}) - \sigma[R_t - E_t(\pi_{t+1})] + u_{1t}, \quad (1)$$

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

$$R_t = \phi_\pi E_t(\pi_{t+1}) + u_{3t}, \quad (3)$$

where y_t , π_t and R_t are output gap, inflation rate, nominal interest rate, respectively, and u_{1t}, u_{2t} and u_{3t} are independent iid standard normal random variables, which respectively represents a shock to the output Euler equation (1), New Keynesian Phillips curve (NKPC) (2), and monetary policy function (3). $E_t(\cdot) = E(\cdot|I_t)$ is the conditional expectation operator conditional on I_t , the information set at time t , σ is the parameter of elasticity of intertemporal substitution, $\delta \in (0, 1)$ is the discount factor, κ is the slope of the NKPC, and ϕ_π controls the reaction of the monetary policy to inflation. Because a solution is

$$\begin{bmatrix} y_t \\ \pi_t \\ R_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & -\sigma \\ \kappa & 1 & -\sigma\kappa \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}, \quad (4)$$

we have covariance restrictions:³

$$Cov([y_t \ \pi_t \ R_t]') = \begin{bmatrix} 1 + \sigma^2 & \kappa + \kappa\sigma^2 & -\sigma \\ \kappa + \sigma^2\kappa & 1 + \kappa^2 + \sigma^2\kappa^2 & -\sigma\kappa \\ -\sigma & -\sigma\kappa & 1 \end{bmatrix}. \quad (5)$$

Suppose that we use

$$f(\sigma, \kappa) = [1 + \sigma^2, \kappa + \sigma^2\kappa, -\sigma, 1 + \kappa^2 + \sigma^2\kappa^2, -\sigma\kappa]', \quad (6)$$

and the corresponding five elements in the covariance matrix of the three observed variables, where we set $\sigma = 1$ and $\kappa = 0.5$. We consider two cases. In case 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

³While there is no unique solution to this model, we simply use a solution from Canova and Sala (2009). This fact does not cause any problem in our minimum distance estimation exercise based on (5).

specified and model B is incorrectly specified. In case 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. Although the two models are both correctly specified in this design, model A is more parsimonious than model B.

Suppose that we employ a classical minimum distance (CMD) estimator and choose the model with a smaller estimation criterion function. Table 1 shows the frequencies of selecting the right model (model A) when one selects a model with a smaller value of the minimized estimation criterion function. The number of Monte Carlo replications is 1,000 and the sample sizes are 50, 100 and 200. The column labeled “Diagonal” indicates the selection probabilities when the diagonal weighting matrix whose diagonal elements are the reciprocals of the bootstrap variances of the sample analogs of the restrictions. The column labeled “Optimal” indicates those when the weighting matrix is the inverse of the bootstrap covariance matrix of the sample analog of the restrictions.

This table shows that although this intuitive procedure tends to select the correctly specified models over the incorrectly specified models in designs 1, it is likely to select an overparameterized model if the two models have equal explanatory power as in population as in cases 2. Our proposed QML model selection criteria overcome this issue as formally shown in the next section.

Furthermore, the issue of weak and partial identification is often encountered in macroeconomic applications. In this static model example, suppose $f(\sigma, \kappa) = [\kappa + \sigma^2\kappa, 1 + \kappa^2 + \sigma^2\kappa^2, -\sigma\kappa]'$ and the corresponding three elements of the covariance matrix are used to estimate the model instead of (6). As κ approaches zero, the strength of identification of σ becomes weaker. In addition, the slope of NKPC κ is known to depend on more than one deep parameters which are partially identified.

Tables A1–A4 in Inoue and Shintani (2017) report the performance of our QML model selection procedure and others when the parameters are strongly, weakly or partially identified in this static model. We find that the probability of selecting model A also tends to approach one as the sample size increases, even when identification is weak and when some parameters are partially identified.

3 Asymptotic Theory

3.1 Quasi-Marginal Likelihood for Extremum Estimators

We first consider QMLs based on general objective functions and establish the consistency of the model selection based on QMLs. Models A and B are parameterized by vectors, $\alpha \in \mathbf{A}$ and $\beta \in \mathbf{B}$, respectively, where $\mathbf{A} \subset \mathfrak{R}^{p_A}$ and $\mathbf{B} \subset \mathfrak{R}^{p_B}$. Let $\hat{q}_{A,T}(\alpha)$ and $\hat{q}_{B,T}(\beta)$ be the objective functions for estimating models A and B, respectively, that would be minimized in the conventional frequentist estimation method. Here subscript T signifies the fact that both objective functions are constructed using the same data with sample size T . The conventional extremum estimators are given by $\hat{\alpha}_T = \arg \min_{\alpha \in \mathbf{A}} \hat{q}_{A,T}(\alpha)$ and $\hat{\beta}_T = \arg \min_{\beta \in \mathbf{B}} \hat{q}_{B,T}(\beta)$. Let $q_A(\alpha)$ and $q_B(\beta)$ denote the population analog of $\hat{q}_{A,T}(\alpha)$ and $\hat{q}_{B,T}(\beta)$, and define the (possibly pseudo) true parameter values of α and β by $\alpha_0 = \arg \min_{\alpha \in \mathbf{A}} q_A(\alpha)$ and $\beta_0 = \arg \min_{\beta \in \mathbf{B}} q_B(\beta)$, respectively.

In this paper, we say model A is nested in model B if there exist a function $\phi : \mathfrak{R}^{p_A} \rightarrow \mathfrak{R}^{p_B}$ such that $q_A(\alpha) = q_B(\phi(\alpha))$, for all $\alpha \in \mathbf{A}$.⁴ Rivers and Vuong (2002) generalize a likelihood ratio test for model comparison originally developed by Vuong (1989) to the case of extremum estimators and 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 $\hat{q}_{A,T}(\hat{\alpha}_T) - \hat{q}_{B,T}(\hat{\beta}_T) = O_p(T^{-\frac{1}{2}})$ otherwise. Hnatkovska, Marmer and Tang (2012) and Smith (1992) show similar results for CMD estimators and GMM estimators, respectively.

Following Chernozhukov and Hong (2003), the quasi-posteriors of models A and B can be defined by

$$\frac{e^{-T\hat{q}_{A,T}(\alpha)}\pi_A(\alpha)}{\int_{\mathbf{A}} e^{-T\hat{q}_{A,T}(\alpha)}\pi_A(\alpha)d\alpha} \text{ and } \frac{e^{-T\hat{q}_{B,T}(\beta)}\pi_B(\beta)}{\int_{\mathbf{B}} e^{-T\hat{q}_{B,T}(\beta)}\pi_B(\beta)d\beta}, \quad (7)$$

where $\pi_A(\alpha)$ and $\pi_B(\beta)$ are the prior probability density functions for the two models. By treating (7) as the posterior, their LTE (e.g., quasi-posterior mean, median and mode) is obtained via Markov Chain Monte Carlo (MCMC) which may be particularly useful when the objective functions are is not numerically tractable or when extremum estimates are not reasonable.

⁴See Vuong(1989) for the formal definition when competing models are nested, strictly non-nested or overlapping.

We now define the QMLs for models A and B by

$$m_A = \int_{\mathbf{A}} e^{-T\hat{q}_{A,T}(\alpha)} \pi_A(\alpha) d\alpha \text{ and } m_B = \int_{\mathbf{B}} e^{-T\hat{q}_{B,T}(\beta)} \pi_B(\beta) d\beta. \quad (8)$$

We say that the QML model selection criterion is consistent in the following sense: $m_A > m_B$ with probability approaching one if $q_A(\alpha_0) < q_B(\beta_0)$ or if $q_A(\alpha_0) = q_B(\beta_0)$ and $p_{A_s} < p_{B_s}$ where p_{A_s} and p_{B_s} are the numbers of strongly identified parameters in models A and B, respectively. For example, suppose the parameter α in model A corresponds to a subset of the parameter β in model B so that two models are nested, and a remaining part of parameter β are fixed at their true value. Then, $\beta_0 = \phi(\alpha_0)$ and $q_A(\alpha_0) = q_B(\beta_0)$ hold and the model with the fixed parameter value is preferable because it reduces parameter estimation uncertainty. This definition of consistency is common in the literature on selection of parametric models (see Leeb and Pötscher, 2009; Nishii, 1988; and Inoue and Kilian, 2006 to name a few), and the model selection criteria, such as those in Nishii (1988), Sin and White (1996) and Hong and Preston (2012), are designed to be consistent in this sense.⁵ In order that the model selection based on the value of $q_A(\alpha_0)$ and p_{A_s} relative to the value of $q_B(\beta_0)$ and p_{B_s} to make sense, every object of model selection needs to be incorporated in criterion functions and parameter vectors. We will be more specific about this issue when we remark on Propositions 1 and 2 in the next two subsections.

It can be shown that the log of the QML is approximated by

$$\ln(m_A) = -T\hat{q}_{A,T}(\hat{\alpha}_T) - \frac{p_A}{2} \ln(T) + O_p(1), \quad (9)$$

$$\ln(m_B) = -T\hat{q}_{B,T}(\hat{\beta}_T) - \frac{p_B}{2} \ln(T) + O_p(1). \quad (10)$$

Because the leading term diverges at rate T , a correctly specified model will be chosen over an incorrectly specified model. When the model is nested and $q_A(\alpha_0) = q_B(\beta_0)$, a more parsimonious model will be chosen because of the second term. The problem is when the model is not nested and $q_A(\alpha_0) = q_B(\beta_0)$. Because the difference of the dominant terms can have either sign with positive probability and diverges at rate $T^{1/2}$, a model will be chosen randomly. When two models that are not nested have equal fit, one may still prefer a more parsimonious model based on Occam's razor or

⁵One could also call a model selection criterion consistent if $m_A > m_B$ with probability approaching one if $q_A(\alpha_0) < q_B(\beta_0)$. Our model selection criterion is also consistent in this sense.

if a selected model is to be used for forecasting (Inoue and Kilian, 2006). For that purpose we propose the following *modified QML*:

$$\ln(\tilde{m}_A) = \ln(m_A) + (T - \sqrt{T})\hat{q}_{A,T}(\hat{\alpha}_T), \quad (11)$$

$$\ln(\tilde{m}_B) = \ln(m_B) + (T - \sqrt{T})\hat{q}_{B,T}(\hat{\beta}_T). \quad (12)$$

In the logarithmic form, the modified QML effectively replaces $-T\hat{q}_{A,T}(\hat{\alpha}_T)$ in the Laplace approximation by $-\sqrt{T}\hat{q}_{A,T}(\hat{\alpha}_T)$ so that $T^{\frac{1}{2}}(\hat{q}_{A,T}(\hat{\alpha}_T) - \hat{q}_{B,T}(\hat{\beta}_T)) = O_p(1)$ and the more parsimonious model will be selected in the equal-fit case, and the modified QML model selection criterion remains consistent for both nested and nonnested models.

Let α_s and β_s denote strongly identified parameters if any, α_w and β_w weakly identified parameters, and α_p and β_p partially identified parameters. p_{A_s} and p_{B_s} denote the number of the strongly identified parameters, p_{A_w} and p_{B_w} the number of the weakly identified parameters, and p_{A_p} and p_{B_p} the number of the partially identified parameters. A_s and B_s are the spaces of the strongly identified parameters, A_w and B_w are those of the weakly identified parameters, and A_p and B_p are those of the partially identified parameters. We consider two cases. Some of the parameters may be weakly identified in the first case ($\alpha = [\alpha'_s \alpha'_w]'$, $p_A = p_{A_s} + p_{A_w}$, and $A = A_s \times A_w$) while some parameters may be partially identified in the second case ($\alpha = [\alpha'_s \alpha'_p]'$, $p_A = p_{A_s} + p_{A_p}$ and $A = A_s \times A_p$). The parameters may be all strongly identified ($p_A = p_{A_s}$), weakly identified ($p_A = p_{A_w}$) or partially identified ($p_A = p_{A_p}$).

Assumption 1

- (a) \mathbf{A} is compact in \mathfrak{R}^{p_A} [\mathbf{B} is compact in \mathfrak{R}^{p_B}].
- (b) If $p_{A_s} > 0$, $\hat{q}_A(\alpha)$ and $q_A(\alpha)$ are twice continuously differentiable in $\alpha_s \in \text{int}(\mathbf{A}_s)$, $\sup_{\alpha \in \mathbf{A}} |\hat{q}_{A,T}(\alpha) - q_A(\alpha)| = o_p(1)$, $\sup_{\alpha \in \mathbf{A}} \|\nabla_{\alpha_s} \hat{q}_{A,T}(\alpha) - \nabla_{\alpha_s} q_A(\alpha)\| = o_p(1)$ and $\sup_{\alpha \in \mathbf{A}} \|\nabla_{\alpha_s}^2 \hat{q}_{A,T}(\alpha) - \nabla_{\alpha_s}^2 q_A(\alpha)\| = o_p(1)$ [If $p_{B_s} > 0$, $\hat{q}_B(\beta)$ and $q_B(\beta)$ are twice continuously differentiable in $\beta_s \in \text{int}(\mathbf{B}_s)$, $\sup_{\beta \in \mathbf{B}} |\hat{q}_{B,T}(\beta) - q_B(\beta)| = o_p(1)$, $\sup_{\beta \in \mathbf{B}} \|\nabla_{\beta_s} \hat{q}_{B,T}(\beta) - \nabla_{\beta_s} q_B(\beta)\| = o_p(1)$ and $\sup_{\beta \in \mathbf{B}} \|\nabla_{\beta_s}^2 \hat{q}_{B,T}(\beta) - \nabla_{\beta_s}^2 q_B(\beta)\| = o_p(1)$].
- (c) If $p_{A_s} > 0$, $\pi_{A_s}(\alpha_s)$ is continuous at $\alpha_{s,0}$ and $\pi_{A_s}(\alpha_{s,0}) > 0$. [If $p_{B_s} > 0$, $\pi_{B_s}(\beta_s)$ is continuous at $\beta_{s,0}$ and $\pi_{B_s}(\beta_{s,0}) > 0$].

Assumption 1(b) requires uniform convergence of $\hat{q}_{A,T}(\cdot)$, $\nabla\hat{q}_{A,T}(\cdot)$ and $\nabla^2\hat{q}_{A,T}(\cdot)$ to $q_A(\cdot)$, $\nabla q_A(\cdot)$ and $\nabla^2 q_A(\cdot)$, respectively, which holds under more primitive assumptions, such as the compactness of the parameter spaces, pointwise convergence and stochastic equicontinuity (see Theorem 1 of Andrews, 1992).

It is well-known that some parameters of DSGE models may not be strongly identified. See Canova and Sala (2009), for example. It is therefore important to investigate asymptotic properties of our model selection procedure in case some parameters may not be strongly identified. To allow for some weakly identified parameters, we impose the following assumptions:

Assumption 2 (Weak Identification)

(a) $q_A(\alpha) = q_{A_s}(\alpha_s) + T^{-1}q_{A_w}(\alpha)$ if $p_{A_s} > 0$ and $q_A(\alpha_w) = T^{-1}q_{A_w}(\alpha_w)$ if $p_{A_s} = 0$ where $q_{A_w}(\cdot)$ is $O_p(1)$ uniformly in $\alpha \in \mathbf{A}$. [$q_B(\beta) = q_{B_s}(\beta_s) + T^{-1}q_{B_w}(\beta)$ if $p_{B_s} > 0$ and $q_B(\beta_w) = T^{-1}q_{B_w}(\beta_w)$ if $p_{B_s} = 0$ where $q_{B_w}(\cdot)$ is $O_p(1)$ uniformly in $\beta \in \mathbf{B}$].

(b) If $p_{A_s} > 0$, then there exists $\alpha_{s,0} \in \text{int}(\mathbf{A}_s)$ such that for every $\epsilon > 0$

$$\inf_{\alpha_s \in \mathbf{A}_s: \|\alpha_s - \alpha_{s,0}\| \geq \epsilon} q_{A_s}(\alpha_s) > q_{A_s}(\alpha_{s,0})$$

[If $p_{B_s} > 0$, then there exists $\beta_{s,0} \in \text{int}(\mathbf{B}_s)$ such that for every $\epsilon > 0$

$$\inf_{\beta_s \in \mathbf{B}_s: \|\beta_s - \beta_{s,0}\| \geq \epsilon} q_{B_s}(\beta_s) > q_{B_s}(\beta_{s,0})].$$

(c) If $p_{A_s} > 0$, the Hessian $\nabla_{\alpha_s}^2 q_{A_s}(\alpha_{s,0})$ is positive definite [If $p_{B_s} > 0$, the Hessian $\nabla_{\beta_s}^2 q_{B_s}(\beta_{s,0})$ is positive definite].

Remarks.

1. Assumptions 1 and 2 are high-level assumptions, and sufficient and lower-level assumptions for CMD and GMM estimators are provided in the next two subsections.
2. Typical prior densities are continuous in macroeconomic applications, and Assumption 1(c) is likely to be satisfied.
3. Assumption 2(a) postulates that α_w is weakly identified while α_s is strongly identified where $\alpha = [\alpha'_s \alpha'_w]'$. Note that we allow for cases in which the parameters are all strongly identified as well as cases in which they are all weakly identified. When

there is a strongly identified parameter, Assumption 2(b) requires that its true parameter value $\alpha_{s,0}$ uniquely minimizes the population estimation criterion function, and Assumption 2(c) requires that the second-order sufficient condition for minimization is satisfied.

Theorem 1 (The Case When Some Parameters May Be Weakly Identified). Suppose that Assumptions 1 and 2 hold.

- (a) If $q_{A_s}(\alpha_{s,0}) < q_{B_s}(\beta_{s,0})$, then $m_A > m_B$ and $\tilde{m}_A > \tilde{m}_B$ with probability approaching one.
- (b) (Nested Case) If $q_{A_s}(\alpha_{s,0}) = q_{B_s}(\beta_{s,0})$, $p_{A_s} < p_{B_s}$ 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 $\tilde{m}_A > \tilde{m}_B$ with probability approaching one.
- (c) (Nonnested Cases) If $q_{A_s}(\alpha_{s,0}) = q_{B_s}(\beta_{s,0})$, $p_{A_s} < p_{B_s}$ and $\hat{q}_{A,T}(\hat{\alpha}_T) - \hat{q}_{B,T}(\hat{\beta}_T) = O_p(T^{-1/2})$, then $\tilde{m}_A > \tilde{m}_B$ with probability approaching one.

Remarks.

1. Theorem 1(a) shows that the proposed QML model selection criterion selects the model with a smaller population estimation 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 fewer number of strongly identified parameters. In the special case where Model A is correctly specified and is a restricted version of Model B, our criterion will select Model A provided the restriction is imposed on a strongly identified parameter. This is because the QML has a built-in penalty term for parameters that are not necessary for reducing the population criterion function as can be seen in the Laplace approximation of the marginal likelihood. Note that these results hold even when the parameters are all strongly identified.

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. In

that case, our model selection criterion will select the better approximating model with probability approaching one.

3. When the models are not nested and $q_A(\alpha_{s,0}) = q_B(\beta_{s,0})$, Theorem 1(c) shows that the marginal likelihood does not necessarily select a more parsimonious model even asymptotically. This is consistent with Hong and Preston's (2012) result on BIC. Although this may not be a major concern when the models are nonnested, the modified QML selects a parsimonious model even when the nonnested models satisfy $q_A(\alpha_{s,0}) = q_B(\beta_{s,0})$. However, because the leading term in the modified QML diverges at rate \sqrt{T} , the modified QML is less powerful than the unmodified QML if $q_A(\alpha_{s,0}) < q_B(\beta_{s,0})$. We will investigate this trade-off in Monte Carlo experiments.

Next we consider cases in which some parameters may be partially identified. We say that the parameters are partially identified if $\mathbf{A}_0 = \{\alpha_0 \in \mathbf{A} : q_A(\alpha_0) = \min_{\alpha \in \mathbf{A}} q_A(\alpha)\}$ consists of more than one points (see Chernozhukov, Hong and Tamer, 2007). Moon and Schorfheide (2012) lists macroeconomic examples in which this type of identification arises. Similarly, we define $\mathbf{B}_0 = \{\beta_0 \in \mathbf{B} : q_B(\beta_0) = \min_{\beta \in \mathbf{B}} q_B(\beta)\}$. In addition to Assumption 1, we impose the following assumptions.

Assumption 3 (Partial Identification)

- (a) There exists $\mathbf{A}_0 \subset \mathbf{A}$ such that, for every $\alpha_0 \in \mathbf{A}_0$ and $\epsilon > 0$ $\inf_{\alpha \in (\mathbf{A}_0^\epsilon)^c} q_A(\alpha) > q_A(\alpha_0)$, where $(\mathbf{A}_0^\epsilon)^c = \{\alpha \in \mathbf{A} : d(\alpha, \mathbf{A}_0) \geq \epsilon\}$ and $d(\alpha, \mathbf{A}_0) = \inf_{a \in \mathbf{A}_0} \|\alpha - a\|$. [There exists $\mathbf{B}_0 \subset \mathbf{B}$ such that, for every $\beta_0 \in \mathbf{B}_0$ and $\epsilon > 0$, $\inf_{\alpha \in (\mathbf{A}_0^\epsilon)^c} q_A(\alpha) > q_A(\alpha_0)$, where $(\mathbf{B}_0^\epsilon)^c = \{\beta \in \mathbf{B} : d(\beta, \mathbf{B}_0) \geq \epsilon\}$ and $d(\beta, \mathbf{B}_0) = \inf_{b \in \mathbf{B}_0} \|\beta - b\|$.]
- (b) If $p_{A_s} > 0$, the Hessian $\nabla_{\alpha_s}^2 q_A([\alpha'_{s,0}, \alpha'_{p,0}]')$ is positive definite for some $\alpha_{p,0} \in \mathbf{A}_{p,0}$. [If $p_{B_s} > 0$, the Hessian $\nabla_{\beta_s}^2 q_B([\beta'_{s,0}, \beta'_{p,0}]')$ is positive definite for some $\beta_{p,0} \in \mathbf{B}_{p,0}$.]
- (c) $\int_{\mathbf{A}_{p,0}} \pi_A(\alpha_p | \alpha_{s,0}) d\alpha_p > 0$ where $\pi_A(\alpha_p | \alpha_s)$ is the prior density of α_p conditional on α_s . [$\int_{\mathbf{B}_{p,0}} \pi_B(\beta_p | \beta_{s,0}) d\beta_p > 0$ where $\pi_B(\beta_p | \beta_s)$ is the prior density of β_p conditional on β_s .]

Remark. Assumptions 3(a), (b) and (c) are generalization of Assumptions 2(b), 2(c)

and 1(c), respectively, to sets.

Theorem 2 (The Case When Some Parameters May Be Partially Identified).

- (a) Suppose that Assumptions 1 and 3(a) hold. If $\min_{\alpha \in \mathbf{A}} q_A(\alpha) < \min_{\beta \in \mathbf{B}} q_B(\beta)$, then $m_A > m_B$ and $\tilde{m}_A > \tilde{m}_B$ with probability approaching one.
- (b) (Nested Case) Suppose that Assumptions 1 and 3 hold. If $q_{A_s}(\alpha_{s,0}) = q_{B_s}(\beta_{s,0})$, $p_{A_s} < p_{B_s}$ 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 $\tilde{m}_A > \tilde{m}_B$ with probability approaching one.
- (c) (Nonnested Cases) Suppose that Assumption 3 holds. If $q_{A_s}(\alpha_{s,0}) = q_{B_s}(\beta_{s,0})$, $p_{A_s} < p_{B_s}$ and $\hat{q}_{A,T}(\hat{\alpha}_T) - \hat{q}_{B,T}(\hat{\beta}_T) = O_p(T^{-1/2})$, then $\tilde{m}_A > \tilde{m}_B$ with probability approaching one.

Remarks

Theorem 3(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.

3.2 Quasi-Marginal Likelihood for CMD Estimators

Since the extremum estimators include some class of important estimators popularly used in practice, it should be useful to describe a set of assumptions specific to each of the estimators. We first consider the CMD estimator, which has been used to estimate the structural parameters in DSGE models by matching the predicted impulse response function (say, DSGE-IRF) and the estimated impulse response function from the VAR models (say, VAR-IRF) in empirical macroeconomics.

Suppose that we compare two DSGE models, models A and B. Models A and B are parameterized by structural parameter vectors, $\alpha \in \mathbf{A}$ and $\beta \in \mathbf{B}$, where $\mathbf{A} \subset \mathfrak{R}^{p_A}$ and $\mathbf{B} \subset \mathfrak{R}^{p_B}$. Let $f(\alpha)$ and $g(\beta)$, of dimension $k \times 1$, denote the DSGE-IRFs of models A and B, respectively. The IRF matching estimator of Christiano, Eichenbaum and Evans (2005) minimizes the criterion functions

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

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

with respect to $\alpha \in \mathbf{A}$ and $\beta \in \mathbf{B}$, respectively, for models A and B, where $\hat{\gamma}_T$ is a $k \times 1$ vector of VAR-IRFs, and \widehat{W}_T is a $k \times k$ positive semidefinite weighting matrix.⁶ It should be noted that the condition for identifying VAR-IRFs must be satisfied in DSGE models. For example, if short-run restrictions are used to identify VAR-IRFs, the restrictions must be satisfied in the DSGE model. Otherwise IRF matching does not yield a consistent estimator and model selection based on IRF matching may become invalid.

Let

$$q_A(\alpha) = \frac{1}{2}(\gamma_0 - f(\alpha))' W (\gamma_0 - f(\alpha)), \quad (13)$$

$$q_B(\beta) = \frac{1}{2}(\gamma_0 - g(\beta))' W (\gamma_0 - g(\beta)), \quad (14)$$

where γ_0 is a vector of population VAR-IRFs and W is a positive definite matrix.

While our model selection depends on the choice of weighting matrices, if one is to calculate standard errors from MCMC draws, \widehat{W}_T needs to be set to the inverse of the asymptotic covariance matrix of $\hat{\gamma}_T$, which eliminates the arbitrariness of the choice of the weighting matrix. When the optimal weighting matrix is not used, the formula in Chernozhukov and Hong (2002) and Kormilitsina and Nekipelov (2016) should be used to calculate standard errors.

For CMD estimators, we make the following assumptions:

Assumption 4.

- (a) $\sqrt{T}(\hat{\gamma}_T - \gamma_0) \xrightarrow{d} N(0_{k \times 1}, \Sigma)$ where Σ is positive definite.
- (b) \mathbf{A} is compact in \mathfrak{R}^{p_A} [\mathbf{B} is compact in \mathfrak{R}^{p_B}].
- (c) $f(\alpha) = f_s(\alpha_s) + T^{-1/2} f_w(\alpha)$ if $p_A > 0$ and $f(\alpha) = T^{-1/2} f_w(\alpha)$ if $p_A = 0$, where $f_s : A_s \rightarrow \mathfrak{R}^k$ and $f_w : A \rightarrow \mathfrak{R}^k$ are three times continuously differentiable in the interior of \mathbf{A} . The Jacobian matrix of f_s , $F_s(\alpha_s) = \partial f_s(\alpha_s) / \partial \alpha'_s$, has rank p_{A_s} at $\alpha_s = \alpha_{s,0}$ [$g : \mathbf{B} \rightarrow \mathfrak{R}^k$ is three times continuously differentiable in the interior of \mathbf{B} , and its Jacobian matrix $G_s(\beta_s) = \partial g(\beta_s) / \partial \beta'_s$ has rank p_{B_s} at $\beta_s = \beta_{s,0}$].

⁶Jordà, 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 $h(\gamma, \alpha) = \gamma - f(\alpha)$.

- (d) \widehat{W}_T is positive semidefinite with probability one and converges in probability to a positive definite matrix W .
- (e) If $p_{A_s} > 0$, there is a unique $\alpha_{s,0} \in \text{int}(\mathbf{A}_s)$ such that $\alpha_{s,0} = \text{argmin}_{\alpha_s \in \mathbf{A}_s} f_s(\alpha_s)' W f_s(\alpha_s)$. If $p_{A_s} = 0$, $f(\alpha) = T^{-1/2} f_w(\alpha_w)$ [If $p_{B_s} > 0$, $g(\beta) = g_s(\beta_s) + T^{-1/2} g_w(\beta)$, and there is a unique $\beta_{s,0} \in \text{int}(\mathbf{B}_s)$ such that $\beta_{s,0} = \text{argmin}_{\beta_s \in \mathbf{B}_s} g_s(\beta_s)' W g_s(\beta_s)$. If $p_{B_s} = 0$, $g(\beta) = T^{-1/2} g_w(\beta_w)$.
- (f) If $p_{A_s} > 0$,

$$F_s(\alpha_{s,0})' W F_s(\alpha_{s,0}) - [(\gamma - f_s(\alpha_{s,0}))' W \otimes I_{p_{A_s,0}}] \frac{\partial \text{vec}(F_s(\alpha_{s,0})')}{\partial \alpha'_{s,0}}$$

is positive definite [If $p_{B_s} > 0$,

$$G_s(\beta_{s,0})' W G_s(\beta_{s,0}) - [(\gamma - g_s(\beta_{s,0}))' W \otimes I_{p_{B_s}}] \frac{\partial \text{vec}(G_s(\beta_{s,0})')}{\partial \beta'_s}$$

is positive definite].

- (g) There is $\mathbf{A}_0 = \{\alpha_0\} \times \mathbf{A}_{p,0} \subset \mathbf{A}$ such that, for any $\alpha \in \mathbf{A}_0$, $(\gamma - f(\alpha))' W (\gamma - f(\alpha)) = \min_{\tilde{\alpha} \in \mathbf{A}} (\gamma - f(\tilde{\alpha}))' W (\gamma - f(\tilde{\alpha})) < (\gamma - f(\bar{\alpha}))' W (\gamma - f(\bar{\alpha}))$ for any $\bar{\alpha} \in \mathbf{A} \cap \mathbf{A}_0^c$ [There is $\mathbf{B}_0 = \{\beta_0\} \times \mathbf{B}_{p,0} \subset \mathbf{B}$ such that, for any $\beta \in \mathbf{B}_0$, $(\gamma - g(\beta))' W (\gamma - g(\beta)) = \min_{\tilde{\beta} \in \mathbf{B}} (\gamma - g(\tilde{\beta}))' W (\gamma - g(\tilde{\beta})) < (\gamma - g(\bar{\beta}))' W (\gamma - g(\bar{\beta}))$ for any $\bar{\beta} \in \mathbf{B} \cap \mathbf{B}_0^c$].

- (h) If $p_{A_s} > 0$, there is $\alpha_{p,0} \in \mathbf{A}_{p,0}$ such that

$$F_s(\alpha)' W F_s(\alpha) - [(\gamma - f(\alpha))' W \otimes I_{p_{A_s}}] \frac{\partial \text{vec}(F_s(\alpha)')}{\partial \alpha'_s},$$

is positive definite at $\alpha = [\alpha'_{s,0} \alpha'_{p,0}]'$ [If $p_{B_s} > 0$, there is $\beta_{p,0} \in \mathbf{B}_{p,0}$ such that

$$G_s(\beta)' W G_s(\beta) - [(\gamma - g(\beta))' W \otimes I_{p_{B_s}}] \frac{\partial \text{vec}(G_s(\beta)')}{\partial \beta'_s},$$

is positive definite at $\beta = [\beta'_{s,0} \beta'_{p,0}]'$.

Remarks.

1. The root T consistency and asymptotic normality of VAR-IRFs follow from stationarity data and restrictions that structural IRFs are point-identified.

2. Assumption 4(e) follows Guerron-Quintana, Inoue and Kilian’s (2013) definition of weak identification in the minimum distance framework.

The model selection based on the QML computed from quasi-Bayesian CMD Estimators is justified by the following proposition.

Proposition 1.

- (a) Under Assumptions 4(a)–(f), Assumptions 1 and 2 hold.
- (b) Under Assumptions 4(a)–(d), (g) and (h), Assumptions 1 and 3 hold.

Remark. In our framework, the VAR-IRF, $\widehat{\gamma}_T$, is common across different DSGE models and our proposed QML is designed to select a DSGE model. When it is used to select VAR-IRFs given a DSGE model, the QML will select all the valid impulse responses, that is, all the VAR-IRFs that equal the corresponding DSGE-IRFs (see Hall, Inoue, Nason and Rossi, 2012, on this issue). When it is used to select the number of lags in the VAR model given an IRF and a DSGE model, the QML will select enough many lags so that the implied VAR-IRF equals the DSGE-IRF provided that the population VAR-IRF equals the corresponding DSGE-IRF for sufficiently many lags. However, it will not necessarily choose the smallest lag for which the implied VAR-IRF equals the DSGE-IRF with probability approaching one. This is because the criterion function is not used in the VAR parameter estimation. If the goal is to select the number of lags, the criterion function and the parameter vector need to be modified.

3.3 Quasi-Marginal Likelihood for GMM Estimators

Another important class of the estimator we consider is the GMM estimator. For GMM estimators, the criterion functions of models A and B are respectively given by

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

$$\hat{q}_{B,T}(\beta) = \frac{1}{2} g_T(\beta)' \widehat{W}_{B,T} g_T(\beta),$$

where $f_T(\alpha) = (1/T) \sum_{t=1}^T f(x_t, \alpha)$, $g_T(\beta) = (1/T) \sum_{t=1}^T g(x_t, \beta)$ and $\widehat{W}_{A,T}$ and $\widehat{W}_{B,T}$ are $k \times k$ positive semidefinite weighting matrices.

Let

$$q_A(\alpha) = \frac{1}{2}E[f(x_t, \alpha)]'W_A E[f(x_t, \alpha)], \quad (15)$$

$$q_B(\beta) = \frac{1}{2}E[g(x_t, \beta)]'W_B E[g(x_t, \beta)], \quad (16)$$

where W_A and W_B are positive definite matrices.

For the GMM estimation, we impose the following assumptions:

Assumption 5.

- (a) $\sup_{\alpha \in \mathbf{A}} \|T^{-1} \sum_{t=1}^T \{f(x_t, \alpha) - E[f(x_t, \alpha)]\}\| = O_p(T^{-1/2})$, $\sup_{\alpha \in \mathbf{A}} \|T^{-1} \sum_{t=1}^T \{(\partial/\partial\alpha')f(x_t, \alpha) - E[(\partial/\partial\alpha')f(x_t, \alpha)]\}\| = o_p(1)$, and $\sup_{\alpha \in \mathbf{A}} \|T^{-1} \sum_{t=1}^T \{(\partial/\partial\alpha')\text{vec}((\partial/\partial\alpha')f(x_t, \alpha) - E[(\partial/\partial\alpha')\text{vec}((\partial/\partial\alpha')f(x_t, \alpha))])\}\| = o_p(1)$ [$\sup_{\beta \in \mathbf{B}} \|T^{-1} \sum_{t=1}^T \{g(x_t, \beta) - E[g(x_t, \beta)]\}\| = O_p(T^{-1/2})$, $\sup_{\beta \in \mathbf{B}} \|T^{-1} \sum_{t=1}^T \{(\partial/\partial\beta')g(x_t, \beta) - E[(\partial/\partial\beta')g(x_t, \beta)]\}\| = o_p(1)$, and $\sup_{\beta \in \mathbf{B}} \|T^{-1} \sum_{t=1}^T \{(\partial/\partial\beta')\text{vec}((\partial/\partial\beta')g(x_t, \beta) - E[(\partial/\partial\beta')\text{vec}((\partial/\partial\beta')g(x_t, \beta))])\}\| = o_p(1)$].
- (b) \mathbf{A} is compact in \mathfrak{R}^{p_A} [\mathbf{B} is compact in \mathfrak{R}^{p_B}].
- (c) $f : \mathcal{X} \times \mathbf{A} \rightarrow \mathfrak{R}^k$ is three times continuously differentiable in α in the interior of \mathbf{A} , and $F_s(\alpha_s) = \partial E[f(x_t, \alpha_s)/\partial\alpha'_s]$ has rank p_{A_s} at $\alpha_s = \alpha_{s,0}$ [$g : \mathcal{X} \times \mathbf{B} \rightarrow \mathfrak{R}^\ell$ is three times continuously differentiable in β in the interior of \mathbf{B} , and $G_s(\beta_s) = E[\partial g(x_t, \beta_s)/\partial\beta'_s]$ has rank p_{B_s} at $\beta_s = \beta_{s,0}$].
- (d) $\widehat{W}_{A,T}$ is positive semidefinite with probability one and converges in probability to a positive definite matrix W_A [$\widehat{W}_{B,T}$ is positive semidefinite with probability one and converges in probability to a positive definite matrix W_B].
- (e) If $p_{A_s} > 0$, $E[f(x_t, \alpha)] = f_s(\alpha_s) + T^{-1/2}f_w(\alpha)$, and there is a unique $\alpha_{s,0} \in \text{int}(\mathbf{A}_s)$ such that $\alpha_{s,0} = \text{argmin}_{\alpha_s \in \mathbf{A}_s} f_s(\alpha_s)'W_A f_s(\alpha_s)$. If $p_{A_s} = 0$, $E[f(x_t, \alpha)] = T^{-1/2}f_w(\alpha_w)$ [If $p_{B_s} > 0$, $E[g(x_t, \beta)] = g_s(\beta_s) + T^{-1/2}g_w(\beta)$, and there is a unique $\beta_{s,0} \in \text{int}(\mathbf{B}_s)$ such that $\beta_{s,0} \in \text{argmin}_{\beta_s \in \mathbf{B}_s} g_s(\beta_s)'W_B g_s(\beta_s)$. If $p_{B_s} = 0$, $E[g(x_t, \beta)] = T^{-1/2}g_w(\beta_w)$].
- (f) If $p_{A_s} > 0$,

$$F_s(\alpha_{s,0})'W_A F_s(\alpha_{s,0}) + [E(f_s(x_t, \alpha_{s,0}))'W_A \otimes I_{p_{A_s}}] \frac{\partial \text{vec}(F_s(\alpha_{s,0})')}{\partial \alpha'_s}$$

is positive definite [If $p_{B_s} > 0$,

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

is positive definite].

- (g) There is $\mathbf{A}_0 = \{\alpha_0\} \times \mathbf{A}_{p,0} \subset \mathbf{A}$ such that, for any $\alpha \in \mathbf{A}_0$, $E[f(x_t, \alpha)]'W_A E[f(x_t, \alpha)] = \min_{\tilde{\alpha} \in \mathbf{A}} E[f(x_t, \tilde{\alpha})]'W_A E[f(x_t, \tilde{\alpha})] < E[f(x_t, \bar{\alpha})]'W_A E[f(x_t, \bar{\alpha})]$ for any $\bar{\alpha} \in \mathbf{A} \cap \mathbf{A}_0^c$ [There is $\mathbf{B}_0 = \{\beta_0\} \times \mathbf{B}_{p,0} \subset \mathbf{B}$ such that, for any $\beta \in \mathbf{B}_0$, $E[g(x_t, \beta)]'W_B E[g(x_t, \beta)] = \min_{\tilde{\beta} \in \mathbf{B}} E[g(x_t, \tilde{\beta})]'W_B E[g(x_t, \tilde{\beta})] < E[g(x_t, \bar{\beta})]'W_B E[g(x_t, \bar{\beta})]$ for any $\bar{\beta} \in \mathbf{B} \cap \mathbf{B}_0^c$].

- (h) If $p_{A_s} > 0$, there is $\alpha_{p,0} \in \mathbf{A}_{p,0}$ such that

$$F_s(\alpha)'W_A F_s(\alpha) + [E(f(x_t, \alpha))'W_A \otimes I_{p_{A_s}}] \frac{\partial \text{vec}(F_s(\alpha)')}{\partial \alpha_s'}$$

is positive definite at $\alpha = [\alpha'_{s,0} \alpha'_{p,0}]'$ [If $p_{B_s} > 0$, there is $\beta_{p,0} \in \mathbf{B}_{p,0}$ such that

$$G_s(\beta)'W_B G_s(\beta) + [E(g(x_t, \beta))'W_B \otimes I_{p_{B_s}}] \frac{\partial \text{vec}(G_s(\beta)')}{\partial \beta_s'}$$

is positive definite at $\beta = [\beta'_{s,0} \beta'_{p,0}]'$.

The model selection based on the QML computed from quasi-Bayesian GMM Estimators is justified by the following proposition.

Proposition 2.

- (a) Under Assumptions 5(a)–(f), Assumptions 1 and 2 hold.
(b) Under Assumptions 5(a)–(d), (g) and (h), Assumptions 1 and 3 hold.

Remark. It can be shown that the Laplace approximation to the QML has a bonus term that is increasing in the number of overidentifying restrictions, i.e., the number of moment conditions minus the number of strongly identified parameters, which is similar to the moment selection criterion of Andrews (1999). Therefore, when our QML criterion is used for selecting moment conditions, it will select as many correctly specified moment conditions as possible as Andrews' (1999) criterion does. While the selected moment conditions are valid, they are not necessarily relevant. If the goal is to select correctly specified and relevant moments, the QML cannot accomplish it alone.

4 Discussions

Although our method is applicable to more general problems, one main motivation for our proposed QML model selection is IRF matching. IRF matching is a limited-information approach and is an alternative to full-information likelihood based approaches, such as MLE and Bayesian approaches. Estimators based on the full-information likelihood are more efficient when the likelihood function is correctly specified, while IRF matching estimators may be more robust to potential misspecification that does not affect the IRF, such as misspecification of distributional forms. In addition to this usual tradeoff between efficiency and robustness, we discuss very particular features IRF matching estimators have.

(a) Bayesian and frequentist inferential frameworks for IRF matching: The use of priors not only alleviates such numerical challenges inherent in IRF matching but also gives a limited-information Bayesian interpretation to IRF matching (Kim, 2002; Zellner, 1998). Because the QML is a function of data, which model is selected is also a function of the data. Because the quasi-posterior distribution is conditional on the data and the result of model selection is a function of the conditioning set, the posterior distribution remains the same. Therefore, there is no issue of post model selection inference in the limited-information Bayesian inferential framework. See Dawid (1994) in the full-information Bayesian inferential framework.

It is often useful to have a frequentist interpretation of Bayesian estimators, especially if the practitioner is not strictly a Bayesian. Chernozhukov and Hong (2002) provides consistency and asymptotic normality of LTE that includes IRF matching estimators as a special case. Our paper shows the consistency of the QML model selection criterion in a frequentist inferential framework. When we interpret inference based on the model selected by our model selection criterion in a frequentist inferential framework, it is likely to suffer from the problem with post model selection inference (Leeb and Pötscher, 2005, 2009), which is typical in model selection.

(b) Identification of structural IRFs: To implement IRF matching, VAR-IRFs and DSGE-IRFs must be identical when the latter is evaluated at the true structural parameter value. Consider two cases. In the first case, the number of structural shocks

in a DSGE model and the number of observed variables are the same. Fernández-Villaverde, Rubio-Ramírez, Sargent and Watson (2007) find a sufficient condition for the two sets of structural IRF to match. One of their conditions is that matrix D in the measurement equation is square and is nonsingular. Even when this condition is not satisfied (e.g., no measurement error), there are cases in which the two structural IRFs coincide. For example, consider

$$x_{t+1} = Ax_t + Bu_{t+1}, \quad (17)$$

$$y_{t+1} = Cx_t, \quad (18)$$

where x_t is a $n \times 1$ vector of state variables, y_t and u_t are $k \times 1$ vectors of observed variables and economic shocks, respectively, and u_t is Gaussian white noise vector with zero mean and covariance matrix I_k . Provided the eigenvalues of A are all less than unity in modulus, y_t has an $MA(\infty)$ representation:

$$y_t = C(I - AL)^{-1}Bu_t = CBu_t + CABu_{t-1} + CA^2Bu_{t-2} + \dots \quad (19)$$

and it is invertible. Thus, the structural IRFs, CB , CAB , CA^2B , ..., can be obtained from a VAR(∞) process together with the short-run restriction that the impact matrix is given by CB . In practice, the VAR(∞) process is approximated by a finite-order VAR(p) model where p is obtained by AIC, for example. In the IRF matching literature it is quite common to build a DSGE model in such a way that CB is lower triangular so that the recursive identification condition can be used to identify structural IRFs from a VAR model. Many DSGE models do not satisfy typical short-run conditions for identifying structural impulse responses. There are at least two approaches. First, one can match IRFs without matching the impact period matrix. Let A_j denote the j th step ahead structural impulse response matrix implied by a DSGE model with A_0 being the impact matrix. Let B_j denote the j th step ahead reduced-form impulse response matrix obtained from a VAR model. Then we have $\Sigma = A_0A_0'$ and $B_jA_0 = A_j$, which in turn can be written as $g(\gamma, \theta) = 0$, where γ is a vector that consists of the elements of B_j 's and the distinct elements of Σ and θ is a vector of DSGE parameters. In the second approach, one can match moments (e.g., Andreasen, Fernández-Villaverde and Rubio-Ramírez, 2016; Kormilitsina and Nekipelov, 2016).

(c) The dimensions of VAR and IRF: The Monte Carlo experiments in Hall,

Inoue, Nason and Rossi (2012) show that the performance of IRF matching estimators deteriorates as the number of IRFs increases. Guerron-Quintana, Inoue and Kilian (2016) show that, when the number of impulse responses is greater than the number of VAR parameters, IRF matching estimators have nonstandard asymptotic distributions because the delta method fails. We conjecture that asymptotic properties of the QML model selection criterion may be affected because the bootstrap covariance matrix estimator is asymptotically singular.

In general, we recommend to select the order of VAR models by information criteria, such as AIC, as done in section 6 because the true VAR representation is likely to be of infinite order. We then suggest to choose the maximum horizon so that the number of impulse response does not exceed the number of VAR parameters to avoid the above issue.

(d) The choice of weighting matrices: The optimal weighting matrix and diagonal weighting matrices are common choices for the weighting matrix. There are two arguments for the optimal weighting matrix. First, when the optimal weighting matrix is used, the IRF matching estimation criterion function can be interpreted as an approximate (log-)likelihood function where the IRF estimate is viewed as “an observation” because the optimal weighting matrix is the inverse of the bootstrap covariance matrix of that observation. Thus it is natural to interpret estimation results from the limited-information Bayesian inferential framework when the optimal weighting matrix is used. Second, when the optimal weighting matrix is used, the generalized information matrix equality of Chernozhukov and Hong (2003) is satisfied. Standard errors can be obtained from standard deviations of MCMC draws in the frequentist inferential framework. When the optimal weighting matrix is not used, one needs to use the sandwich formula in Chernozhukov and Hong (2003) or bootstrap the entire MCMC algorithm to obtain correct standard errors. The main argument for diagonal weighting matrices is based on computational tractability of the resulting estimation criterion function. Even if a proposal density is poorly chosen because of the numerical behavior of the estimation criterion function based on the optimal weighting matrix, MCMC draws should still converge to the quasi posterior distribution although it may require a larger number of draws. Third, one can multiply the estimation criterion

function by any number without changing its optimum. Using the optimal weighting matrix eliminates such arbitrariness in the CMD and GMM frameworks.

(e) The use of modified QMLs: There may be some cases the modified QMLs are recommended over the (unmodified) QMLs in model selection because of the possible inconsistency. One possibility is the case of point mass mixture priors which include a mass at a point mixed with a continuous distribution. For example, suppose two alternative models of nominal exchange rates, S_t , as

Model A (IMA(1) model) $\alpha = \theta$: $\Delta S_t = \varepsilon_t + \theta \varepsilon_{t-1}$.

Model B (AR(2) model) $\beta = (\phi_1, \phi_2)$: $S_t = \phi_1 S_{t-1} + \phi_2 S_{t-2} + \varepsilon_t$.

The two models are non-nested but are equivalent under the random walk specification, namely, if $\theta = 0$ in model A and $(\phi_1, \phi_2) = (1, 0)$ in model B. Because the random walk model is known to be supported by many previous empirical studies as a preferred model for the nominal exchange rate, it make sense to employ point mass priors at $\theta = 0$ in model A and $(\phi_1, \phi_2) = (1, 0)$ in model B. If the true model is the random walk model, the (unmodified) QML will select model B with positive probability. The modified QML will select model A over model B with probability approaching one because the former is more parsimonious, however.

5 Guide for practitioners

We describe how to implement our procedure in this section. First, we specify a quasi-likelihood function and estimate the model by the random-walk Metropolis Hasting algorithm. As discussed in the previous section, we recommend to use the optimal weighting matrix, the inverse of the covariance matrix for bootstrap IRF estimates although we also consider the diagonal weighting matrix in the Monte Carlo experiments in section 6. As suggested in An and Schorfheide (2007), we set the proposal distribution to $N(\alpha^{(j-1)}, c\hat{H}^{-1})$ where $\alpha^{(0)} = \hat{\alpha}$, $c = 0.3$ 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. The draw α from $N(\alpha^{(j-1)}, c[\nabla^2 \hat{q}_{A,T}(\hat{\alpha})]^{-1})$ is accepted with probability

$$\min \left(1, \frac{e^{-T\hat{q}_{A,T}(\alpha)} \pi_A(\alpha)}{e^{-T\hat{q}_{A,T}(\alpha^{(j-1)})} \pi_A(\alpha^{(j-1)})} \right). \quad (20)$$

In section 6, we use 50,000 draws.

Second, we compute the QML using the last half of the draws. We consider four methods for computing the QML: Laplace approximations, modified harmonic estimators of Geweke (1998) and Sims, Waggoner, and Zha (2008) and the estimator of Chib and Jeliazkov (2001). For the Laplace approximation, we evaluate the QML by

$$e^{-Tq_A(\hat{\alpha})} \left(\frac{T}{2\pi} \right)^{\frac{k-p_A}{2}} \pi_A(\hat{\alpha}) |\widehat{W}_T|^{\frac{1}{2}} |\nabla^2 \hat{q}_{A,T}(\hat{\alpha})|^{-\frac{1}{2}}, \quad (21)$$

at the quasi-posterior mode, $\hat{\alpha}$ (here subscript T is omitted for notational simplicity). In our Monte Carlo experiment, we use 20 randomly chosen starting values for a numerical optimization routine to obtain the posterior mode. We use 1,000 bootstrap replications to obtain the bootstrap covariance matrix of IRF estimators in the Monte Carlo experiments.

In the modified harmonic mean method, the QML is computed as the reciprocal of

$$E \left[\frac{w(\alpha)}{\exp(-T\hat{q}_T(\alpha)) \pi_A(\alpha)} \right], \quad (22)$$

which is evaluated using MCMC draws, given a weighting function $w(\alpha)$. We consider two alternative choice of weighting functions which have been proposed in the literature. The first choice is suggested by Geweke (1999), who sets $w(\alpha)$ to be the truncated normal density

$$w(\alpha) = \frac{\exp[-(\alpha - \tilde{\alpha})' \tilde{V}_\alpha^{-1} (\alpha - \tilde{\alpha})/2] \mathbf{1}\{(\alpha - \tilde{\alpha})' \tilde{V}_\alpha^{-1} (\alpha - \tilde{\alpha}) < \chi_{p_A, \tau}^2\}}{(2\pi)^{p_A/2} |\tilde{V}_\alpha|^{1/2} \tau},$$

where $\tilde{\alpha}$ is the quasi-posterior mean, \tilde{V}_α is the quasi-posterior covariance matrix, $\mathbf{1}\{\cdot\}$ is an indicator function, $\chi_{p_A, \tau}^2$ is the 100 τ th percentile of the chi-square distribution with p_A degrees of freedom, and $\tau \in (0, 1)$ is a constant. The second choice is the one proposed by Sims, Waggoner, and Zha (2008). They point out that Geweke's (1999) method may not work well when the posterior distribution is non-elliptical, and suggest an weighting function given by

$$w(\alpha) = \frac{\Gamma(p_A/2)}{2\pi^{p_A/2} |\hat{V}_\alpha|^{1/2}} \frac{f(r)}{r^{p_A-1}} \frac{\mathbf{1}\{-T\hat{q}(\alpha) + \ln \pi_A(\alpha) > L_{1-q}\}}{\bar{\tau}},$$

where \hat{V}_α is the second moment matrix centered around the quasi-posterior mode $\tilde{\alpha}$, $f(r) = [vr^{v-1}/(c_{90}^v/0.9 - c_1^v)] \mathbf{1}\{c_1 < r < c_{90}/(0.9)^{1/v}\}$, $v = \ln(1/9)/\ln(c_{10}/c_{90})$, $r = [(\alpha - \hat{\alpha})' \hat{V}_\alpha^{-1} (\alpha - \hat{\alpha})]^{1/2}$, c_j is the j th percentile of the distance r , L_{1-q} is the 100(1- q)

percentile of the log quasi-posterior distribution, $q \in (0, 1)$ is a constant, and $\bar{\tau}$ is the quasi-posterior mean of $\mathbf{1}\{-T\hat{q}(\alpha) + \ln \pi_A(\alpha) > L_{1-q}\}\mathbf{1}\{c_1 < r < c_{90}/(0.9)^{1/v}\}$. Following Herbst and Schorfheide (2015), we consider $\tau = 0.5$ and 0.9 in the estimator of Geweke (1999) and $q = 0.5$ and 0.9 in the estimator of Sims, Waggoner, and Zha (2008). In the Monte Carlo and empirical application sections, we only report the results for $\tau = 0.9$ and $q = 0.9$ to save space.⁷

For the estimator of Chib and Jeliazkov (2001), the log of the QML is evaluated by

$$\ln \pi_A(\tilde{\alpha}) - T\hat{q}_{A,T}(\tilde{\alpha}) - \ln \hat{p}_A(\tilde{\alpha}) \quad (23)$$

where

$$\hat{p}_A(\alpha) = \frac{(1/J) \sum_{j=1}^J r(\alpha^{(j)}, \tilde{\alpha}) \phi_{\tilde{\alpha}, c^2 \tilde{\Sigma}}(\alpha^{(j)})}{(1/K) \sum_{k=1}^K r(\tilde{\alpha}, \alpha^{(k)})}, \quad (24)$$

$\phi_{\tilde{\alpha}, c^2 \tilde{\Sigma}}(\cdot)$ is the pdf of $N(\tilde{\alpha}, c^2 \tilde{\Sigma})$ and $r(\tilde{\alpha}, \alpha^{(k)})$ is the acceptance probability of moving $\tilde{\alpha}$ to $\alpha^{(k)}$ in the Metropolis Hasting algorithm. The numerator of (24) is evaluated using the last 50% 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 25,000 so that $K = J$. $c^2 \tilde{\Sigma}$ is either set to the one used in the proposal density or estimated from the posterior draws.

The modified QML (11) requires minimizing the estimation criterion function, which may defeat the purpose of using the quasi-Bayesian approach. Instead we approximate it by averaging the values of the log of the quasi-posterior density over quasi-posterior draws:

$$E[\ln(\pi_A(\alpha)) - T\hat{q}_{A,T}(\alpha)], \quad (25)$$

where the expectation is with respect to the quasi-posterior draws. This is computationally tractable because it can be calculated from MCMC draws. Because the quasi-posterior distribution will concentrate around $\hat{\alpha}$ asymptotically, and the log prior is $O(1)$ and does not affect the divergence rate of the modified QML, the resulting modified QML model selection criterion remains consistent as analyzed in the previous section. Our Monte Carlo results show that this approximation works well.

⁷The Monte Carlo results for $\tau = 0.5$ and $q = 0.5$ are reported in Tables A5–A9 in Inoue and Shintani (2017).

6 Monte Carlo Experiments

We investigate the small-sample properties of the QML using the small-scale DSGE model considered in Guerron-Quintana, Inoue and Kilian (2016) that consists of

$$y_t = E(y_{t+1}|I_{t-1}) - \sigma [E(R_t|I_{t-1}) - E(\pi_{t+1}|I_{t-1}) - z_t], \quad (26)$$

$$\pi_t = \delta E(\pi_{t+1}|I_{t-1}) + \kappa y_t, \quad (27)$$

$$R_t = \rho_r R_{t-1} + (1 - \rho_r)(\phi_\pi \pi_t + \phi_y y_t) + \xi_t, \quad (28)$$

where y_t , π_t and R_t denote the output gap, inflation rate, and nominal interest rate, respectively, and I_t denotes the information set at time t . The technology and monetary policy shocks follow

$$z_t = \rho_z z_{t-1} + \sigma^z \varepsilon_t^z, \quad (29)$$

$$\xi_t = \sigma^r \varepsilon_t^r, \quad (30)$$

where ε_t^z and ε_t^r are independent iid standard normal random variables. Note that the timing of the information is nonstandard, e.g., $E(\pi_{t+1}|I_{t-1})$ instead of $E(\pi_{t+1}|I_t)$ in the NKPC. The idea behind these information restrictions is to capture that the economy reacts slowly to a monetary policy shock while it reacts contemporaneously to technology shocks. Specifically, inflation does not contemporaneously reacts to monetary policy shocks but it does to technology shocks in this model. We impose such recursive short-run restrictions to identify VAR-IRFs. In the data generating process, we set $\kappa = 0.025$, $\sigma = 1$, $\delta = 0.99$, $\phi_\pi = 1.5$, $\phi_y = 0.125$, $\rho_r = 0.75$, $\rho_z = 0.90$, $\sigma^z = 0.30$, $\sigma^r = 0.20$ as in Guerron-Quintana, Inoue and Kilian (2016).

We consider four cases. In cases 1 and 3, κ , σ^{-1} and ρ_r are estimated in model A, and κ and ρ_r are estimated in model B with $\sigma^{-1} = 3$. The other parameters are set to the true parameter values. In cases 2 and 4, σ^{-1} and ρ_r are estimated in model A with κ set to its true parameter value and κ , σ^{-1} and ρ_r are estimated in model B. In other words, model B is misspecified in cases 1 and 3. and model A is more parsimonious than model B in cases 2 and 4.

We use a bivariate VAR(p) model of inflation and the nominal interest rate to estimate structural impulse responses. To identify structural impulse responses, we use the short-run restriction that inflation does not respond to the monetary policy

shock contemporaneously, which is satisfied in the above model. In cases 1 and 2, all the structural impulse responses up to horizon H are used in LTE. In cases 3 and 4, only the structural impulse responses to the technology shock (up to horizon H) are used. We use the AIC to select the VAR lag order where p is selected from $\{H, H + 1, \dots, [5(T/\ln(T))^{0.25}]\}$ where $[x]$ is the integer part of x . We set the lower bound on p to H . When p is smaller than H , the asymptotic distribution of VAR-IRFs is singular and our theoretical results do not hold. See Guerron-Quintana, Inoue and Kilian (2016) on inference on VAR-IRFs in such cases.

We consider $T = 50, 100, 200$ and $H = 2, 4, 8$. The number of Monte Carlo simulations is set to 1,000, the number of random-walk Metropolis-Hasting draws is 50,000, the number of bootstrap draws for computing the weighting matrix is 1,000.

Tables 2 and 3 report the probabilities of selecting model A in cases 1 and 2 and those in cases 3 and 4, respectively. The lower parts of Tables 2 and 3 show that the method of selecting a model based on the value of the estimation criterion function performs poorly when two models are both correctly specified and one model is more parsimonious than the other. The tables show that the probabilities of the QML's selecting the right model tend to increase as the sample size grows. As conjectured in section 3, the QML performs better than the modified QML when one model is correctly specified and the other is misspecified, and the modified QML outperforms the QML when both are correctly specified and one model is more parsimonious than the other. Using a fewer IRFs, that is, using the IRFs to the technology shock only, improves the performance of the QML. These tables show that the different methods for computing the QML do not produce a substantial or systematic difference in the performance in large samples. The diagonal weighting matrix provides better performance than the optimal weighting matrix but the difference becomes smaller as the sample size grows.⁸

To shed light on the accuracy of QML estimates further, we report the means and standard deviations of 100 QML estimates from a realization of data in Table 4. Except when many impulse responses are used when the sample size is small (the third row in the table), the standard deviations appear reasonably small. Furthermore, the differences across the methods are small.

⁸As shown in Tables A5–A8 in Inoue and Shintani (2017), the results are not sensitive to the choice of the turning parameters q and τ .

7 Empirical Applications

7.1 New Keynesian Phillips Curve: GMM Estimation

In this section, we apply our procedure to choose between alternative specifications of the structural Phillips curve under nonzero trend inflation when the models are estimated using quasi-Bayesian GMM. Let $\hat{\pi}_t = \pi_t - \pi$ be the log-deviation of aggregate inflation π_t from the trend inflation π , and $\hat{ulc}_t = ulc_t - ulc$ be the log-deviation of unit labor cost ulc_t from its steady-state ulc .

In Galí and Gertler (1999), the hybrid New Keynesian Phillips Curve (hereafter NKPC) is derived from a Calvo (1983) type staggered price setting model with firms set prices using indexation to trend inflation π with probability ξ_p (see also Yun, 1996). For the remaining $1 - \xi_p$ fraction of the firms, $1 - \omega$ fraction of firms set prices optimally but the remaining ω fraction are rule-of-thumb (ROT) price setters who set their prices equal to the average price set in the most recent round of price adjustments with a correction based on the lagged inflation rate. Under these conditions, a hybrid NKPC can be derived as

$$\hat{\pi}_t = \gamma_b \hat{\pi}_{t-1} + \gamma_f E_t \hat{\pi}_{t+1} + \kappa \hat{ulc}_t \quad (31)$$

with its coefficients given by

$$\gamma_b = \frac{\omega}{\xi_p + \omega[1 - \xi_p(1 - \delta)]}, \gamma_f = \frac{\delta \xi_p}{\xi_p + \omega[1 - \xi_p(1 - \delta)]}, \kappa = \frac{(1 - \xi_p)(1 - \delta \xi_p)(1 - \omega)}{\xi_p + \omega[1 - \xi_p(1 - \delta)]},$$

where $\delta \in (0, 1)$ is a discounted factor.

In Smets and Wouters (2003, 2007), a partial indexation specification is used instead of the ROT specification of Galí and Gertler (1999). In their specification, firms set prices at an optimal level with probability $1 - \xi_p$. For the remaining ξ_p fraction of the firms, prices are determined as a weighted sum of lagged inflation and trend inflation (or steady state inflation) with an weight ι_p on the lagged inflation. Under these conditions, an alternative hybrid NKPC can be derived as (31) with coefficients given by

$$\gamma_b = \frac{\iota_p}{1 + \iota_p \delta}, \gamma_f = \frac{\delta}{1 + \iota_p \delta}, \kappa = \frac{(1 - \xi_p)(1 - \delta \xi_p)}{\xi_p(1 + \iota_p \delta)},$$

where $\iota_p \in [0, 1]$ is the degree of partial indexation to lagged inflation. Note that when $\omega = 0$ in the ROT specification and $\iota_p = 0$ in the partial indexation specification, both

hybrid NKPCs become the baseline NKPC with only forward looking firms ($\gamma_b = 0$ and $\gamma_f = \delta$).

In the previous empirical literature, the classical GMM has often been employed to estimate the hybrid NKPC. In our quasi-Bayesian GMM estimation, we utilize the orthogonality condition of expectation error to past information, as well as the definition of $\hat{\pi}_t = \pi_t - \pi$, and estimate structural parameters $\alpha = [\xi_p, \omega, \pi]'$ for the first model and $\beta = [\xi_p, \iota_p, \pi]'$ for the second model. In particular, for the first model, the objective function is $\hat{q}_{A,T}(\alpha) = (1/2)f_T(\alpha)' \widehat{W}_{A,T} f_T(\alpha)$ where $f_T(\alpha) = (1/T) \sum_{t=1}^T f(x_t, \alpha)$, $f(x_t, \alpha) = [\mathbf{z}_t' u_t, \hat{\pi}_t]'$,

$$u_t = \hat{\pi}_t - \gamma_b \hat{\pi}_{t-1} - \gamma_f \hat{\pi}_{t+1} - \kappa ulc_t$$

and \mathbf{z}_t is a vector of instruments. The objective function for the second model can be similarly defined. Optimal weighting matrix \widehat{W} is computed from the HAC estimator with the Bartlett kernel and Andrews' (1991) automatic bandwidth.

For the estimation, we use US quarterly data of the inflation rate based on the GDP implicit price deflator for π_t and the labor income share in the non-farm business sector for ulc_t . As for the choice of instruments \mathbf{z}_t , we follow Galí and Gertler (1999): four lags of inflation, labor income share, long-short interest rate spread, output gap, wage inflation, and commodity price inflation. We use the same set of instruments so that the number of moment conditions is the same for the two NKPCs. For the sample periods, we consider the Great Inflation period (from 1966:Q1 to 1982:Q3) and Post-Great Inflation period (from 1982:Q4 to 2016:Q4). δ is fixed at 0.99.

The list of the structural parameters in our analysis, quasi-Bayesian estimates and prior distributions are reported in Table 5 and the posterior distributions are shown in Figure 1 for both ROT specification and partial indexation specification. The prior and posterior means tend to differ which may suggest that the parameters are strongly identified in these models. The trend inflation rate became substantially lower after the Great Inflation period, as expected. The slope of the Phillips curve (κ) got flattened in the post Great Inflation period compared to the Great Inflation period mainly due to the increased degree of price stickiness (ξ_p). The figure also shows that the slope of the Phillips curve became more spread. However, in general, both estimates of structural and reduced form parameters differs between two specification. ⁹

⁹While the number of parameters is the same between the two models, the joint restrictions on the range

Table 6 reports the QMLs for the two specifications, along with the value of the estimation criterion function and Andrews’(1999) criterion. The results based on comparing QMLs suggest that the ROT specification of Galí and Gertler (1999) outperforms the partial indexation specification of Smets and Wouters (2003, 2007) for both sample periods we consider. In particular, according to Jeffreys’(1961) terminology, the former model is decisively better than the latter model.¹⁰ For the value of the estimation criterion function and Andrews’(1999) criterion, the model with a smaller value should be selected. When these alternative methods are employed, the ROT specification are selected for the first subsample as in the case of QMLs, but conflicting results are obtained for the second subsample. However, since our Monte Carlo results suggest that QMLs are more accurate than the value of the estimation criterion function for selecting correctly specified models in moderate sample sizes, the ROT specification is likely to be the better-fitting specification in the second subsample.

7.2 The Medium-Scale DSGE Model: IRF Matching Estimation

As a second empirical application of our procedure, we consider quasi-Bayesian IRF matching estimation of the medium-scale DSGE Model. For the purpose of evaluating the relative importance of various frictions in the model estimated by the standard Bayesian method, Smets and Wouters (2007) utilize 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,

of parameters are different. For example, the ratio of the forward-looking parameter and backward-looking parameter (γ_f/γ_b) for the ROT specification depends on three parameters (ξ_p, ω, δ) while the ratio for the partial indexation specification depends only on two parameters (ι_p, δ). Such a tighter restriction for the latter model can make the difference in the empirical performance of two models.

¹⁰Provided the prior probabilities are equal, the difference in the QML is *decisive* and *very strong* according to Jeffreys (1961, p.433) and Kass and Raftery (1995, p.777), respectively.

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 QMLs based on the standard DSGE model estimated by Christiano, Trabandt and Walentin (2011). Based on an estimated VAR(2) model of 14 variables using the US quarterly data from 1951Q1 to 2008Q4, they employ short-run and long-run identifying restrictions to compute IRF to (i) a monetary policy shock, (ii) a neutral technology shock and (iii) an investment-specific technology shock. The model is then 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).¹¹ The list of estimated structural parameters in our analysis, quasi-Bayesian estimates and the prior distribution, are reported in Table 7. This estimated model serves as the baseline model when we compare with other models using QMLs.

Following Smets and Wouters (2007), the sources of frictions of the baseline model are divided 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, and 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 8. 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. Let us first consider the role of nominal frictions. According to Jeffreys'(1961) terminology, QMLs are decisively reduced when the degree of nominal price and wage stickiness (ξ_p and ξ_w) is set at 0.10. In contrast, even if the price and wage indexation parameters (ι_p and ι_w) are set at a very small value of 0.01, the values of QMLs are quite similar to that of the baseline model. Thus, we can conclude that Calvo-type

¹¹In our analysis, both price markup and wage markup parameters are fixed at 1.2.

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. 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.

8 Concluding Remarks

In this paper we establish the consistency of the model selection criterion based on the QML obtained from Laplace-type estimators. We consider cases in which parameters are strongly identified and are weakly identified. Our Monte Carlo results confirmed our consistency results. Our proposed procedure is also applied to select an appropriate specification in New Keynesian macroeconomic models using US data.

Our proposed model selection criterion is useful when one selects a model, estimates the structural parameters of the selected model and interpret them. While Bayesian model averaging will select the correct model asymptotically, weights are nonzero in finite samples. It is not clear how to interpret structural parameters of different DSGE models that are estimated simultaneously. Bayesian model averaging may be more useful for forecasting. Application of Bayesian model averaging to IRF matching is beyond the scope of this paper and is left for future research.

References

- An, Sungbae, and Frank Schorfheide (2007), “Bayesian Analysis of DSGE Models,” *Econometric Reviews*, 26, 113–172.
- Andreasen, Martin M., Jesús Fernández-Villaverde and Juan Rubio-Ramírez (2016), “The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications,” Working Paper.
- Andrews, Donald W.K., (1991), “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, 59 (3), 817–858.
- Andrews, Donald W.K., (1992), “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.
- Calvo, Guillermo A. (1983), “Staggered Prices in a Utility-Maximizing Framework,” *Journal of Monetary Economics*, 12 (3), 383–398.
- 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 Jeliazkov (2001), “Marginal Likelihood from the Metropolis-Hastings Output,” *Journal of the American Statistical Association*, 96, 270–281.
- 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 (2016), “Unemployment and Business Cycles,” *Econometrica*, 84, 1523–1569.

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.

Dawid, A.P. (1994), “Selection Paradoxes of Bayesian Inference,” in *Multivariate Analysis and its Applications*, Volume 24, eds., T.W. Anderson, K. A.-T. A. Fang and L. Olkin, IMS: Philadelphia, PA.

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.

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

Fernández-Villaverde, Jesús, Juan F. Rubio-Ramírez, Thomas J. Sargent, and Mark W. Watson (2007), “ABCs (and Ds) of Understanding VARs,” *American Economic Reviews*, 97, 1021–1026.

Fernández-Villaverde, Jesús, Juan F. Rubio-Ramírez, and Frank Schorfheide (2016), “Solution and Estimation Methods for DSGE Models,” in *Handbook of Macroeconomics*, Volume 2A, eds., J.B. Taylor and H. Uhlig, North-Holland, 527–724.

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

Gemma, Yasufumi, Takushi Kurozumi and Mototsugu Shintani (2017), “Trend Inflation and Evolving Inflation Dynamics: A Bayesian GMM Analysis of the

Generalized New Keynesian Phillips Curve,” Bank of Japan IMES Discussion Paper Series No. 2017-E-10.

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* 4, 197–229.

Guerron-Quintana, Pablo, Atsushi Inoue and Lutz Kilian (2016), “Impulse Response Matching Estimators for DSGE Models,” accepted for publication in *Journal of Econometrics*.

Hall, Alastair R., Atsushi Inoue, James M. Nason and Barbara Rossi (2012), “Information Criteria for Impulse Response Function Matching Estimation of DSGE Models,” *Journal of Econometrics*, 170, 499–518.

Herbst, Edward P., and Frank Schorfheide (2015), *Bayesian Estimation of DSGE Models*. Princeton, NJ: Princeton University Press.

Hnatkovska, Victoria, Vadim Marmer and Yao Tang (2012), “Comparison of Misspecified Calibrated Models: The Minimum Distance Approach,” *Journal of Econometrics*, 169, 131–138.

Hong, Han and Bruce Preston (2012), “Bayesian Averaging, Prediction and Nonnested Model Selection,” *Journal of Econometrics*, 167, 358–369.

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

Inoue, Atsushi, and Mototsugu Shintani (2017), “Online Appendix to “Quasi-Bayesian Model Selection,” unpublished manuscript, Vanderbilt University and University of Tokyo.

Jeffreys, Harold (1961), *Theory of Probability*, Third Edition, Oxford University Press.

Jordà, Òscar, and Sharon Kozicki (2011), “Estimation and Inference by the Method of Projection Minimum Distance,” *International Economic Review*, 52,

461–487.

Kass, Robert E., and Adrian E. Raftery (1995), “Bayes Factors,” *Journal of the American Statistical Association*, 430, 773–795.

Kim, Jae-Young (2002), “Limited Information Likelihood and Bayesian Analysis,” *Journal of Econometrics*, 107 , 175–193.

Kim, Jae-Young (2014), “An Alternative Quasi Likelihood Approach, Bayesian Analysis and Data-based Inference for Model Specification, ” *Journal of Econometrics*, 178 , 132–145.

Kormilitsina, Anna, and Denis Nekipelov (2016), “Consistent Variance of the Laplace Type Estimators: Application to DSGE Models,” *International Economic Reviews*, 57, 603–622.

Leeb, Hannes, and Benedikt M. Pötscher (2005), “Model Selection and Inference: Facts and Fiction,” *Econometric Theory*, 21, 21–59.

Leeb, Hannes, and Benedikt M. Pötscher (2009), “Model Selection,” in T.G. Andersen, R.A. Davis, J.-P. Kreiss and T. Mikosch eds., *Handbook of Financial Time Series*, Springer-Verlag.

Miyamoto, Wataru, and Thuy Lan Nguyen (2017), “Understanding the Cross-country Effects of U.S. Technology Shocks, ” *Journal of International Economics*, 106, 143–164.

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

Nishii, R. (1988), “Maximum Likelihood Principle and Model Selection when the True Model is Unspecified,” *Journal of Multivariate Analysis*, 27, 392–403.

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

Rivers, Douglas, and Quang Vuong (2002), “Model Selection Tests for Nonlinear Dynamic Models,” *Econometrics Journal*, 5, 1-39.

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

- Shin, Minchul (2014), “Bayesian GMM,” unpublished manuscript, University of Pennsylvania.
- Sims, Christopher A., Daniel F. Waggoner and Tao Zha (2008), “Methods for Inference in Large Multiple-Equation Markov-Switching Models,” *Journal of Econometrics*, 146, 255–274.
- 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 (2003), “An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area,” *Journal of the European Economic Association*, 1 (5), 1123–1175.
- Smets, Frank, and Rafael Wouters (2007), “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach,” *American Economic Review*, 97, 586–606.
- Smith, Richard J. (1992), “Non-Nested Tests for Competing Models Estimated by Generalized Method of Moments,” *Econometrica*, 60(4), 973–980.
- 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.
- Yun, Tack (1996), “Nominal Price Rigidity, Money Supply Endogeneity, and Business Cycles,” *Journal of Monetary Economics*, 37 (2-3), 345–370.
- Zellner, Arnold (1998), “Past and Recent Results on Maximal Data Information Priors,” *Journal of Statistical Research*, 32 (1), 1–22.

Table 1: Frequencies of selecting model A when the estimation criterion function alone is used

Design	T	Diagonal	Optimal
1	50	1.000	1.000
	100	1.000	1.000
	200	1.000	1.000
2	50	0.000	0.003
	100	0.000	0.000
	200	0.000	0.000

Notes. The restrictions

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

and the corresponding elements of the covariance matrix are used. In design 1, Model A is correctly specified while Model B is misspecified. In design 2, Models A and B are both correctly specified and Model A is more parsimonious than Model B. T denotes the sample size. “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.

Table 2: Frequencies of selecting model A when all impulse responses are used

T	H	Weight Matrix	\hat{q}_T	QML				Modified QML		
				Laplace	Geweke	SWZ	CJ	Geweke	SWZ	CJ
Case 1: Model A is correctly specified and Model B is misspecified										
50	2	Diag	0.77	0.99	0.98	0.99	0.97	0.63	0.96	0.70
		Opt	0.69	0.82	0.82	0.85	0.82	0.31	0.63	0.34
50	4	Diag	0.78	0.99	0.99	0.99	0.98	0.67	0.96	0.75
		Opt	0.72	0.73	0.79	0.76	0.77	0.33	0.58	0.34
50	8	Diag	0.74	0.96	0.97	0.97	0.96	0.63	0.86	0.66
		Opt	0.54	0.25	0.45	0.33	0.38	0.30	0.27	0.27
100	2	Diag	0.70	0.99	0.99	0.99	0.97	0.87	0.99	0.91
		Opt	0.69	0.90	0.92	0.91	0.88	0.39	0.81	0.44
100	4	Diag	0.78	0.99	0.99	0.99	0.98	0.90	0.98	0.93
		Opt	0.73	0.89	0.93	0.90	0.89	0.51	0.85	0.58
100	8	Diag	0.85	0.99	0.99	0.99	0.99	0.92	0.99	0.95
		Opt	0.68	0.81	0.87	0.84	0.84	0.50	0.75	0.50
200	2	Diag	0.71	0.99	0.99	0.99	0.99	0.99	0.99	0.99
		Opt	0.62	0.93	0.94	0.93	0.91	0.50	0.86	0.56
200	4	Diag	0.73	0.99	0.99	0.99	0.99	0.99	0.99	0.99
		Opt	0.72	0.94	0.96	0.95	0.93	0.69	0.91	0.76
200	8	Diag	0.80	0.99	0.99	0.99	0.99	0.98	0.99	0.99
		Opt	0.75	0.91	0.94	0.92	0.91	0.64	0.86	0.71
Case 2: Model A is more parsimonious than model Model B										
50	2	Diag	0.05	0.99	0.97	0.93	0.95	0.99	0.99	0.99
		Opt	0.27	0.92	0.87	0.88	0.88	0.94	0.93	0.94
50	4	Diag	0.05	0.99	0.96	0.94	0.96	0.99	0.99	0.99
		Opt	0.33	0.90	0.85	0.85	0.85	0.90	0.91	0.91
50	8	Diag	0.07	0.98	0.95	0.94	0.94	0.96	0.97	0.96
		Opt	0.41	0.76	0.65	0.68	0.65	0.75	0.79	0.76
100	2	Diag	0.05	0.99	0.97	0.95	0.97	0.99	0.99	0.99
		Opt	0.24	0.92	0.89	0.89	0.90	0.94	0.92	0.93
100	4	Diag	0.04	0.99	0.97	0.96	0.95	0.99	0.99	0.99
		Opt	0.36	0.91	0.90	0.90	0.88	0.94	0.92	0.94
100	8	Diag	0.06	0.99	0.97	0.95	0.95	0.99	0.99	0.99
		Opt	0.38	0.84	0.83	0.84	0.83	0.89	0.87	0.88
200	2	Diag	0.03	0.99	0.96	0.95	0.95	0.99	0.99	0.99
		Opt	0.25	0.88	0.86	0.85	0.87	0.92	0.88	0.92
200	4	Diag	0.04	0.99	0.97	0.96	0.97	0.99	0.99	0.99
		Opt	0.35	0.92	0.91	0.90	0.92	0.94	0.92	0.93
200	8	Diag	0.04	0.99	0.96	0.96	0.96	0.99	0.99	0.99
		Opt	0.38	0.88	0.85	0.85	0.86	0.90	0.88	0.89

Notes: T denotes the sample size and H denotes the maximum horizon for impulse responses. “Diag” refers to cases in which the weighting matrix is diagonal and their diagonal elements are the reciprocals of the bootstrap variances of impulse responses. “Opt” refers to cases in which the weighting matrix is set to the inverse of the bootstrap covariance matrix of impulse responses. \hat{q}_T refers to the method that chooses the model whose estimated criterion function is smaller. “Laplace”, “Geweke”, “SWZ” and “CJ” refer to Laplace approximations, Geweke’s (1998) modified harmonic estimator, Sims, Waggoner and Zha’s (2008) estimator and the estimator of Chib and Jeliazkov (2001), respectively. The numbers in the table are the actual probabilities of selecting Model A over Model B over 1,000 Monte Carlo iterations.

Table 3: Frequencies of selecting model A when only impulse responses to the technology shock are used

T	H	Weight Matrix	\hat{q}_T	QML				Modified QML		
				Laplace	Geweke	SWZ	CJ	Geweke	SWZ	CJ
Case 3: Model A is correctly specified and Model B is misspecified										
50	2	Diag	0.98	0.99	0.98	0.99	0.97	0.61	0.96	0.65
		Opt	0.97	0.99	0.97	0.99	0.96	0.41	0.86	0.46
50	4	Diag	0.98	0.99	0.97	0.99	0.95	0.57	0.95	0.60
		Opt	0.90	0.90	0.92	0.92	0.90	0.43	0.83	0.48
50	8	Diag	0.93	0.98	0.97	0.98	0.96	0.60	0.92	0.62
		Opt	0.62	0.50	0.62	0.57	0.58	0.30	0.50	0.27
100	2	Diag	0.98	0.99	0.99	0.99	0.95	0.74	0.99	0.82
		Opt	0.99	0.99	0.99	0.99	0.96	0.67	0.98	0.74
100	4	Diag	0.99	0.99	0.99	0.99	0.97	0.76	0.99	0.86
		Opt	0.97	0.99	0.99	0.99	0.97	0.69	0.98	0.76
100	8	Diag	0.98	0.99	0.99	0.99	0.98	0.80	0.99	0.87
		Opt	0.82	0.85	0.88	0.87	0.88	0.61	0.82	0.67
200	2	Diag	0.99	0.99	0.99	0.99	0.97	0.92	0.99	0.95
		Opt	0.99	0.99	0.99	0.99	0.97	0.87	0.99	0.93
200	4	Diag	0.98	0.99	0.99	0.99	0.95	0.89	0.99	0.92
		Opt	0.97	0.99	0.99	0.99	0.97	0.87	0.99	0.93
200	8	Diag	0.99	0.99	0.99	0.99	0.97	0.91	0.99	0.95
		Opt	0.84	0.96	0.98	0.95	0.95	0.85	0.89	0.87
Case 4: Model A is more parsimonious than model Model B										
50	2	Diag	0.03	0.99	0.97	0.95	0.96	0.99	0.99	0.99
		Opt	0.07	0.99	0.96	0.95	0.96	0.99	0.99	0.99
50	4	Diag	0.04	0.98	0.96	0.94	0.95	0.99	0.99	0.99
		Opt	0.32	0.99	0.97	0.95	0.96	0.98	0.99	0.98
50	8	Diag	0.06	0.98	0.96	0.95	0.96	0.99	0.98	0.98
		Opt	0.48	0.87	0.82	0.80	0.80	0.84	0.86	0.86
100	2	Diag	0.06	0.97	0.95	0.93	0.95	0.99	0.99	0.99
		Opt	0.07	0.99	0.97	0.96	0.96	0.99	0.99	0.99
100	4	Diag	0.06	0.98	0.95	0.95	0.95	0.99	0.99	0.99
		Opt	0.20	0.99	0.97	0.97	0.96	0.99	0.99	0.99
100	8	Diag	0.05	0.99	0.97	0.95	0.96	0.99	0.99	0.99
		Opt	0.58	0.97	0.96	0.95	0.94	0.96	0.98	0.97
200	2	Diag	0.04	0.95	0.94	0.93	0.92	0.99	0.99	0.99
		Opt	0.09	0.99	0.97	0.97	0.97	0.99	0.99	0.99
200	4	Diag	0.08	0.96	0.94	0.92	0.93	0.99	0.99	0.99
		Opt	0.18	0.99	0.98	0.98	0.97	0.99	0.99	0.99
200	8	Diag	0.07	0.96	0.93	0.93	0.93	0.99	0.99	0.99
		Opt	0.58	0.99	0.97	0.97	0.96	0.98	0.99	0.99

Notes: See the notes for Table 2.

Table 4: The mean and standard deviation of the QML estimates

T	H	Geweke		SWZ				CJ			
		$\tau = 0.5$	$\tau = 0.9$	$q = 0.5$		$q = 0.9$					
All impulse responses are used											
50	2	-44.51	[0.75]	-44.27	[0.73]	-41.83	[0.68]	-42.29	[0.76]	-41.60	[0.68]
50	4	-72.61	[5.00]	-72.13	[5.09]	-73.99	[2.53]	-73.77	[2.55]	-69.76	[4.14]
50	8	-148.88	[11.85]	-148.41	[11.88]	-150.75	[11.23]	-150.70	[11.25]	-146.60	[11.85]
100	2	-85.68	[0.76]	-85.46	[0.80]	-83.01	[0.67]	-83.48	[0.80]	-82.74	[0.65]
100	4	-81.17	[0.78]	-80.97	[0.83]	-78.55	[0.67]	-79.04	[0.80]	-78.28	[0.64]
100	8	-85.30	[0.81]	-85.12	[0.86]	-82.72	[0.68]	-83.20	[0.83]	-82.49	[0.64]
200	2	-113.45	[0.80]	-113.26	[0.87]	-110.86	[0.66]	-111.35	[0.81]	-110.59	[0.64]
200	4	-85.42	[0.78]	-85.30	[0.88]	-82.99	[0.60]	-83.45	[0.74]	-82.79	[0.63]
200	8	-102.28	[0.81]	-102.15	[0.88]	-99.83	[0.63]	-100.29	[0.78]	-99.61	[0.63]
Only impulse responses to the technology shock are used											
50	2	-24.07	[0.77]	-23.84	[0.74]	-21.37	[0.68]	-21.79	[0.77]	-21.16	[0.65]
50	4	-39.71	[0.82]	-39.53	[0.79]	-37.21	[0.67]	-37.55	[0.77]	-36.98	[0.66]
50	8	-58.66	[0.69]	-58.47	[0.67]	-56.23	[0.70]	-56.53	[0.75]	-56.10	[0.72]
100	2	-27.70	[0.70]	-27.51	[0.70]	-25.06	[0.68]	-25.51	[0.78]	-24.85	[0.64]
100	4	-27.94	[0.72]	-27.76	[0.74]	-25.33	[0.68]	-25.79	[0.78]	-25.10	[0.63]
100	8	-29.19	[0.73]	-29.01	[0.74]	-26.58	[0.68]	-27.03	[0.78]	-26.35	[0.63]
200	2	-31.40	[0.71]	-31.27	[0.77]	-28.90	[0.66]	-29.31	[0.74]	-28.76	[0.64]
200	4	-30.54	[0.72]	-30.41	[0.78]	-28.08	[0.63]	-28.50	[0.72]	-27.90	[0.64]
200	8	-34.65	[0.73]	-34.52	[0.80]	-32.22	[0.64]	-32.64	[0.73]	-32.02	[0.66]

Notes: The means and standard deviations (SD) in each row are calculated from 100 QML estimates given a realization of data. See the notes to table 2.

Table 5: Prior and posteriors of parameters of hybrid NKPCs

Parameter	Prior				Quasi-posterior			
	Dist	Mean	Std	Great Inflation		Post Great Inflation		
				Mean	[5%, 95%]	Mean	[5%, 95%]	
<i>(a) ROT specification (Galí and Gertler, 1999)</i>								
Trend inflation	π	Norm	3.50	1.50	5.91	[5.37, 6.45]	2.24	[2.06, 2.42]
Price stickiness	ξ_p	Beta	0.50	0.10	0.68	[0.61, 0.76]	0.88	[0.83, 0.91]
ROT fraction	ω	Beta	0.50	0.10	0.56	[0.47, 0.65]	0.52	[0.38, 0.67]
Backward-looking	$\gamma_{b,A}$	-	0.50	0.07	0.45	[0.40, 0.51]	0.37	[0.30, 0.44]
Forward-looking	$\gamma_{f,A}$	-	0.50	0.07	0.55	[0.49, 0.60]	0.63	[0.56, 0.70]
Slope of NKPC	κ_A	-	0.14	0.09	0.037	[0.019, 0.056]	0.006	[0.003, 0.009]
<i>(b) Partial indexation specification (Smets and Wouters, 2003, 2007)</i>								
Trend inflation	π	Norm	3.50	1.50	5.82	[5.28, 6.37]	2.25	[2.07, 2.43]
Price stickiness	ξ_p	Beta	0.50	0.10	0.79	[0.75, 0.84]	0.91	[0.89, 0.93]
Price indexation	ι_p	Beta	0.50	0.10	0.65	[0.54, 0.75]	0.48	[0.35, 0.62]
Backward-looking	$\gamma_{b,B}$	-	0.33	0.05	0.39	[0.35, 0.43]	0.32	[0.26, 0.38]
Forward-looking	$\gamma_{f,B}$	-	0.67	0.04	0.60	[0.57, 0.65]	0.67	[0.61, 0.74]
Slope of NKPC	κ_B	-	0.40	0.26	0.036	[0.020, 0.053]	0.006	[0.003, 0.009]

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

Table 6: QML estimates of hybrid NKPCs

	QML				\hat{q}_T	Andrews
	Laplace	Geweke	SWZ	CJ		
Great Inflation Period						
(a) <i>ROT</i>	-15.6	-15.3	-12.7	-17.2	0.0597	-88.7
(b) <i>Partial indexation</i>	-20.3	-20.0	-17.4	-21.9	0.0603	-88.6
Post Great Inflation Period						
(a) <i>ROT</i>	-30.2	-28.8	-26.1	-31.7	0.0490	-99.7
(b) <i>Partial indexation</i>	-33.4	-32.1	-29.4	-35.0	0.0484	-99.9

Note: “Laplace”, “Geweke”, “SWZ” and “CJ” refer to Laplace approximations, Geweke’s (1998) modified harmonic estimator, Sims, Waggoner and Zha’s (2008) estimator and the estimator of Chib and Jeliazkov (2001), respectively. \hat{q}_T refers to the estimated criterion function and “Andrews” refers to Andrews’ (1999) model selection criterion.

Table 7: Prior and posteriors of parameters of the baseline DSGE model

Parameter	Prior			Quasi-posterior		
	Dist.	Mean	Std	Mean	[5%, 95%]	
<i>Price-setting rule</i>						
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]
<i>Monetary policy rule</i>						
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]
<i>Preference and technology</i>						
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]
<i>Shocks</i>						
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.

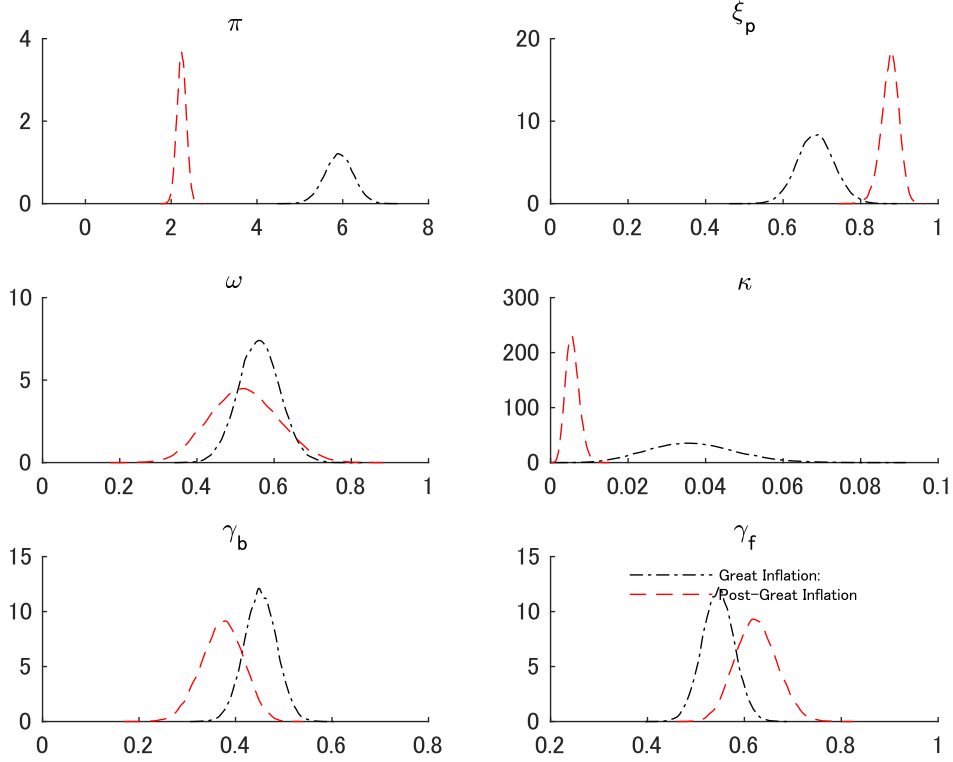
Table 8: Empirical importance of the nominal and real frictions

	Nominal frictions					Real frictions		
	Base	$\xi_p=0.1$	$\xi_w=0.1$	$\iota_p=0.01$	$\iota_w=0.01$	$S''=2$	$b=0.1$	$\sigma_a=0.1$
<i>Quasi-marginal likelihood</i>								
Laplace	370	341	146	369	373	327	279	366
Geweke	366	340	143	368	371	326	276	364
<i>Quasi-posterior mean</i>								
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 the Laplace approximation (Laplace approx.) and the modified harmonic mean estimator of Geweke. SW denotes marginal likelihood estimates from Smets and Wouters (2007).

Figure 1: Posterior distribution of hybrid NKPCs

(a) ROT specification



(b) Partial indexation specification

