Supplement to "Quasi-Bayesian model selection"

(*Quantitative Economics*, Vol. 9, No. 3, November 2018, 1265–1297)

Atsushi Inoue
Department of Economics, Vanderbilt University

Mototsugu Shintani
Faculty of Economics and RCAST, The University of Tokyo

A1. Monte Carlo experiments

In this section, we present full descriptions and results of the Monte Carlo experiments in Inoue and Shintani (2018).

A1.1 *The static model*

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

$$y_t = E_t(y_{t+1}) - \sigma\big(R_t - E_t(\pi_{t+1})\big) + u_{1t}, \tag{A1}$$

$$\pi_t = \delta E_t(\pi_{t+1}) + \kappa y_t + u_{2t}, \tag{A2}$$

$$R_t = \phi_\pi E_t(\pi_{t+1}) + u_{3t}, \tag{A3}$$

where y_t, π_t, and R_t are output gap, inflation rate and nominal interest rate, respectively, and u_{1t}, u_{2t}, and u_{3t} are independent i.i.d. standard normal random variables, which respectively represents a shock to the output Euler equation (A1), New Keynesian Phillips curve (NKPC) (A2), and monetary policy function (A3). $E_t(\cdot) = E(\cdot|I_t)$ is the conditional expectation operator conditional of 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

$$\begin{bmatrix} y_t \\ \pi_t \\ R_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & -\sigma \\ \kappa & 1 & -\sigma\kappa \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}, \tag{A4}$$

Atsushi Inoue: atsushi.inoue@vanderbilt.edu
Mototsugu Shintani: moto.shintani@gmail.com

Available at http://qeconomics.org. https://doi.org/10.3982/QE587

we have covariance restrictions:[1]

$$\mathrm{Cov}\big([y_t\ \pi_t\ R_t]'\big) = \begin{bmatrix} 1+\sigma^2 & \kappa+\kappa\sigma^2 & -\sigma \\ \kappa+\sigma^2\kappa & 1+\kappa^2+\sigma^2\kappa^2 & -\sigma\kappa \\ -\sigma & -\sigma\kappa & 1 \end{bmatrix}. \tag{A5}$$

We consider six cases in each of which two models, model A and model B, are compared. In the first two cases, case 1 and case 2, suppose that we use

$$f(\sigma,\kappa) = \big[1+\sigma^2, \kappa+\sigma^2\kappa, -\sigma, 1+\kappa^2+\sigma^2\kappa^2, -\sigma\kappa\big]', \tag{A6}$$

and the corresponding five elements in the covariance matrix of the three observed variables, where we set $\sigma = 1$ and $\kappa = 0.5$. In these two cases, the parameters are globally and locally identified. 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 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 case, model A is more parsimonious than model B.

In the next two cases, case 3 and case 4, we use

$$f(\sigma,\kappa) = \big[\kappa+\sigma^2\kappa, 1+\kappa^2+\sigma^2\kappa^2, -\sigma\kappa\big]' \tag{A7}$$

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

In the last two cases, cases 5 and 6, the parameters are partially identified in that we estimate ξ_p and θ and the restrictions depend on them only through $\kappa = (1-\xi_p)(1-0.99\xi_p)\theta/\xi_p$ (note that δ is set at 0.99). We use the five restrictions (A6) and set $\sigma = 1$, $\xi_p = 0.5$ and $\theta = 1$ so that $\kappa \approx 0.5$ as in case 1. In case 5, two parameters, ξ_p and θ, are estimated in model A, while the value of σ is set to the correct value, 1, in model A and is set to an incorrect value, 0.5, in model B. In case 6, only ξ_p and θ are estimated while the value of σ is set to the true value in model A, whereas the three parameters are all estimated in model B.

Note that in each of the six cases, model A is always preferred to model B because model A is correctly specified in cases 1, 3, and 5 and is more parsimonious in cases 2, 4, and 6. Table A1 summarizes the six cases as well as the parameter values used in the Monte Carlo simulation experiments.

We estimate the quasi-marginal likelihood (QML) of each model by four methods: the Laplace approximation, the modified harmonic mean estimator of Geweke (1999), Sims, Waggoner, and Zha's (2008) modified harmonic mean estimator, and the estimator

[1]While 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 (A5).

TABLE A1. Simulation design.

		True Model		Model A		Model B	
Identification	Case	σ	κ	σ	κ	σ	κ
Strong	1	1	0.5	estimated	estimated	estimated	fixed at 1
	2	1	0.5	estimated	fixed at 0.5	estimated	estimated
Weak	3	1	0.5	estimated	estimated	estimated	fixed at 1
	4	1	0.5	estimated	fixed at 0.5	estimated	estimated
Partial	5	1	$\xi_p = 0.5$ $\theta = 1$	fixed at 1	estimated (ξ_p, θ)	fixed at 0.5	estimated (ξ_p, θ)
	6	1	$\xi_p = 0.5$ $\theta = 1$	fixed at 1	estimated (ξ_p, θ)	estimated	estimated (ξ_p, θ)

Note: In the cases of strong and partial identification (cases 1, 2, 5 and 6),

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

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

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

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

of Chib and Jeliazkov (2001). For each of the four estimators, we consider the diagonal weighting matrix and the optimal weighting matrix and our proposed modified version of the QML. In addition, we also consider Hong and Preston's (2012) model selection criterion. The number of Monte Carlo replications is set to 1000, the number of randomly chosen initial values for numerical optimization is set to 20, and the sample sizes are 50, 100, and 200. The flat prior is used for all the parameters, and the number of MCMC draws is set to 100,000.

Tables A2 and A3 report the frequencies of selecting model A. In case 1, our proposed criteria tend to select the correctly specified model (model A) over the incorrectly specified model (model B) regardless of the methods used to estimate the QML. The probabilities of selecting model A based on the modified QML are smaller than those based on the unmodified QML. This is because the modification halves the divergence rate of the QML in these cases. In case 2, the frequencies of choosing the more parsimonious model (model A) based on the QML are not as high as those in case 1 because the consistency depends on logarithmic rates as opposed to linear rates in terms of the log QML. The modified QML performs better than the unmodified QML in this case. Because the models are both correctly specified, the less parsimonious model always has a smaller estimation criterion value, the leading term in Laplace approximations, in finite samples. Our modification halves such effects and makes the logarithmic divergence rate of the second term more effective.

The probability of selecting model A also tends to approach one as the sample size increases, even when identification is weak (cases 3 and 4) and when some parameters are partially identified (cases 5 and 6). The proposed modified QML outperforms the Hong and Preston (2012) criterion in the first three cases and in the fourth case for suffi-

TABLE A2. Frequencies of selecting model A (static model).

				QML Based on the RWMH Algorithm					Modified QML Based on the RWMH Algorithm				
				Geweke		SWZ			Geweke		SWZ		
T	Weight Matrix	HP	QML Laplace	$\tau=0.5$	$\tau=0.9$	$q=0.5$	$q=0.9$	CJ	$\tau=0.5$	$\tau=0.9$	$q=0.5$	$q=0.9$	CJ
						Case 1							
50	Diagonal	0.54	1.00	1.00	1.00	1.00	1.00	1.00	0.85	0.85	0.84	0.85	0.84
	Optimal		1.00	1.00	1.00	1.00	1.00	1.00	0.91	0.91	0.89	0.90	0.90
100	Diagonal	0.67	1.00	1.00	1.00	1.00	1.00	1.00	0.88	0.88	0.88	0.88	0.88
	Optimal		1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.96	0.95	0.95	0.96
200	Diagonal	0.77	1.00	1.00	1.00	1.00	1.00	1.00	0.95	0.95	0.95	0.95	0.95
	Optimal		1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.99
						Case 2							
50	Diagonal	0.88	0.85	0.85	0.85	0.85	0.85	0.85	1.00	1.00	1.00	1.00	1.00
	Optimal		0.84	0.83	0.83	0.84	0.84	0.84	0.99	0.99	1.00	1.00	1.00
100	Diagonal	0.93	0.91	0.91	0.91	0.91	0.91	0.91	1.00	1.00	1.00	1.00	1.00
	Optimal		0.92	0.92	0.92	0.92	0.92	0.92	1.00	1.00	1.00	1.00	1.00
200	Diagonal	0.95	0.94	0.94	0.94	0.94	0.94	0.94	1.00	1.00	1.00	1.00	1.00
	Optimal		0.95	0.95	0.95	0.95	0.95	0.95	1.00	1.00	1.00	1.00	1.00
						Case 3							
50	Diagonal	0.56	0.95	0.95	0.95	0.93	0.94	0.94	0.97	0.97	0.95	0.97	0.97
	Optimal		0.96	0.96	0.97	0.95	0.96	0.96	0.96	0.96	0.94	0.95	0.95
100	Diagonal	0.76	0.98	0.98	0.98	0.98	0.98	0.98	0.96	0.96	0.94	0.95	0.95
	Optimal		0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.97	0.96	0.97	0.97
200	Diagonal	0.82	0.99	0.99	0.99	0.99	0.99	0.99	0.93	0.93	0.91	0.92	0.92
	Optimal		0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.96	0.95	0.96	0.96

Note: "HP," "Laplace," "Geweke," "SWZ," and "CJ" refer to Hong and Preston (2012) criterion, Laplace approximation, Geweke's (1999) modified harmonic mean estimator, Sims, Waggoner, and Zha's (2008) modified harmonic mean estimator and the estimator of Chib and Jeliazkov (2001), respectively. τ and q are the truncation parameters for Geweke's (1999) and Sims, Waggoner, and Zha's (2008) methods. See Table A1 for descriptions of the simulation design. "Diagonal" refers to cases in which the weighting matrix is diagonal and their diagonal elements are the reciprocals of the bootstrap variances of the sample analog of the restrictions. "Optimal" refers to cases in which the weighting matrix is set to the inverse of the bootstrap covariance matrix of the sample analog of the restrictions. The numbers in the table are the frequencies of selecting Model A over Model B over 1000 Monte Carlo iterations.

ciently large sample sizes, which might be due to numerical issues with maximizing the empirical likelihood. The Hong and Preston (2012) criterion cannot be computed in the last two cases in which the parameters are partially identified.

Finally, we investigate the accuracy of the QML estimates. Table A4 reports the means and standard deviations of 100 QML estimates given a realization of data as well as the value of QML obtained from numerical integration. The results are based on the optimal weighting matrix. In the first four cases (model A in cases 1 to 4), the differences among the means in each row are very small and are very close to the values based on numerical integration. In the last two partially identified cases (models A and B in case 5), the QML estimates are biased in the same direction. Even though individual QML estimates are biased, model A is selected over model B in case 5 because the sign of the differences in QMLs can be less biased.

TABLE A3. Frequencies of selecting model A (static model).

				QML Based on the RWMH Algorithm					Modified QML Based on the RWMH Algorithm				
				Geweke		SWZ			Geweke		SWZ		
T	Weight Matrix	HP	QML Laplace	$\tau=0.5$	$\tau=0.9$	$q=0.5$	$q=0.9$	CJ	$\tau=0.5$	$\tau=0.9$	$q=0.5$	$q=0.9$	CJ
						Case 4							
50	Diagonal	0.92	0.76	0.75	0.71	0.90	0.87	0.84	0.73	0.65	0.98	0.97	0.91
	Optimal		0.74	0.67	0.64	0.74	0.73	0.72	0.80	0.76	0.92	0.90	0.89
100	Diagonal	0.94	0.88	0.87	0.86	0.94	0.92	0.91	0.94	0.90	1.00	1.00	0.98
	Optimal		0.84	0.82	0.81	0.84	0.83	0.83	0.97	0.96	1.00	1.00	0.99
200	Diagonal	0.97	0.94	0.94	0.94	0.96	0.95	0.95	0.99	0.99	1.00	1.00	1.00
	Optimal		0.87	0.86	0.86	0.87	0.86	0.86	1.00	1.00	1.00	1.00	1.00
						Case 5							
50	Diagonal	NA	0.81	0.77	0.78	0.74	0.73	0.79	0.58	0.58	0.57	0.56	0.59
	Optimal		0.65	0.64	0.64	0.62	0.61	0.65	0.44	0.44	0.42	0.42	0.45
100	Diagonal	NA	0.86	0.89	0.89	0.82	0.82	0.90	0.65	0.65	0.60	0.61	0.66
	Optimal		0.78	0.76	0.75	0.73	0.73	0.77	0.43	0.43	0.44	0.43	0.43
200	Diagonal	NA	0.93	0.97	0.97	0.91	0.91	0.97	0.71	0.70	0.67	0.67	0.71
	Optimal		0.90	0.91	0.91	0.85	0.85	0.93	0.49	0.49	0.51	0.51	0.51
						Case 6							
50	Diagonal	NA	0.54	0.61	0.61	0.53	0.53	0.61	0.69	0.69	0.60	0.61	0.69
	Optimal		0.54	0.55	0.56	0.52	0.51	0.57	0.60	0.60	0.54	0.54	0.60
100	Diagonal	NA	0.64	0.60	0.60	0.54	0.54	0.60	0.67	0.68	0.61	0.62	0.68
	Optimal		0.63	0.63	0.63	0.57	0.57	0.63	0.65	0.65	0.59	0.60	0.65
200	Diagonal	NA	0.63	0.66	0.67	0.60	0.60	0.67	0.71	0.72	0.65	0.66	0.72
	Optimal		0.69	0.65	0.66	0.61	0.60	0.66	0.68	0.68	0.62	0.63	0.68

Note: See the notes for Table A2.

A1.2 *The small-scale DSGE model*

Next, we use the small-scale DSGE model considered in Guerron-Quintana, Inoue, and Kilian (2017) that consists of

$$y_t = E(y_{t+1}|I_{t-1}) - \sigma\big[E(R_t|I_{t-1}) - E(\pi_{t+1}|I_{t-1}) - z_t\big], \tag{A8}$$

$$\pi_t = \delta E(\pi_{t+1}|I_{t-1}) + \kappa y_t, \tag{A9}$$

$$R_t = \rho_r R_{t-1} + (1-\rho_r)(\phi_\pi \pi_t + \phi_y y_t) + \xi_t, \tag{A10}$$

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

$$z_t = \rho_z z_{t-1} + \sigma_z \varepsilon_t^z, \tag{A11}$$

$$\xi_t = \sigma_r \varepsilon_t^r, \tag{A12}$$

where ε_t^z and ε_t^r are mutually independent i.i.d. standard normal random variables. Note that the timing of the information is nonstandard, for example, $E(\pi_{t+1}|I_{t-1})$ instead of

TABLE A4. The mean and standard deviation of QML estimates.

		Geweke				SWZ					
		$\tau=0.5$		$\tau=0.9$		$q=0.5$		$q=0.9$		CJ	
T	True	Mean	[Std]	Mean	[Std]	Mean	[Std]	Mean	[Std]	Mean	[Std]
						Case 1, Model A					
50	−3.218	−3.219	[0.016]	−3.220	[0.008]	−3.363	[0.023]	−3.307	[0.013]	−3.290	[0.011]
100	−5.220	−5.222	[0.018]	−5.222	[0.007]	−5.304	[0.024]	−5.276	[0.013]	−5.261	[0.012]
200	−3.830	−3.830	[0.016]	−3.831	[0.006]	−3.865	[0.021]	−3.853	[0.011]	−3.845	[0.011]
						Case 2, Model A					
50	−1.652	−1.653	[0.015]	−1.653	[0.005]	−1.706	[0.021]	−1.693	[0.008]	−1.676	[0.007]
100	−3.988	−3.987	[0.018]	−3.987	[0.005]	−4.036	[0.021]	−4.022	[0.008]	−4.007	[0.008]
200	−1.653	−1.653	[0.016]	−1.653	[0.006]	−1.665	[0.018]	−1.662	[0.008]	−1.659	[0.007]
						Case 3, Model A					
50	−1.998	−1.998	[0.018]	−1.997	[0.019]	−2.105	[0.024]	−2.060	[0.016]	−2.056	[0.014]
100	−2.703	−2.704	[0.016]	−2.704	[0.006]	−2.738	[0.022]	−2.719	[0.011]	−2.720	[0.011]
200	−2.582	−2.582	[0.017]	−2.583	[0.005]	−2.600	[0.020]	−2.590	[0.010]	−2.590	[0.011]
						Case 4, Model A					
50	−1.079	−1.079	[0.016]	−1.080	[0.005]	−1.138	[0.020]	−1.124	[0.009]	−1.105	[0.007]
100	−1.522	−1.523	[0.015]	−1.521	[0.005]	−1.546	[0.017]	−1.537	[0.008]	−1.530	[0.007]
200	−1.191	−1.195	[0.016]	−1.191	[0.005]	−1.212	[0.020]	−1.205	[0.008]	−1.201	[0.007]
						Case 5, Model A					
50	−2.547	−6.411	[0.458]	−6.302	[0.424]	−6.192	[0.380]	−6.207	[0.387]	−6.250	[0.423]
100	−4.381	−8.237	[0.489]	−8.139	[0.435]	−8.033	[0.402]	−8.044	[0.397]	−8.088	[0.436]
200	−2.019	−8.485	[1.150]	−8.402	[1.127]	−8.347	[1.081]	−8.398	[1.123]	−8.377	[1.087]
						Case 5, Model B					
50	−4.683	−8.672	[0.445]	−8.571	[0.404]	−8.924	[0.404]	−8.772	[0.399]	−8.665	[0.483]
100	−9.321	−13.145	[0.447]	−13.050	[0.410]	−12.946	[0.382]	−12.962	[0.384]	−12.998	[0.410]
200	−12.757	−21.395	[0.425]	−21.306	[0.384]	−21.277	[0.351]	−21.309	[0.351]	−21.349	[0.384]

Note: The means and standard deviations in each row are calculated from 100 QML estimates given a realization of data. See the notes for Table A2.

$E(\pi_{t+1}|I_t)$ in the NKPC (A9). The idea behind these information restrictions is to capture the fact that the economy reacts slowly to a monetary policy shock while it reacts contemporaneously to technology shocks. Specifically, inflation does not contemporaneously react to monetary policy shocks but it does, in this model, to technology shocks. 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 (2017).

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 contemporaneously respond to the monetary policy shock, 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 $\{H, H+1, \ldots, [5(T/\ln(T))^{0.25}]\}$ where $[x]$ is the integer part of x. We set the lower bound on p to H. When p is smaller than H, the variance-covariance matrix of the asymptotic distribution of VAR-IRFs is singular and our theoretical results do not hold. See Guerron-Quintana, Inoue, and Kilian (2017) for the results in such cases.

We consider $T = 50, 100, 200$, and $H = 2, 4, 8$. The number of Monte Carlo simulations is set to 1000, the number of random-walk Metropolis–Hastings draws is 50,000, the number of bootstrap draws for computing the weighting matrix is 1000.

Tables A5, A6, A7, and A8 report the probabilities of selecting the right model in cases 1, 2, 3, and 4, respectively. As in the static model, Tables A6 and A8 show that the model selection method 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 of Inoue and Shintani (2018), 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 based on the optimal weighting matrix. 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. The results are not sensitive to the choice of the tuning parameters q and τ.

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 A9. Except for the cases with the longer horizon and small sample size (the second and third rows in the table), the standard deviations appear reasonably small. Furthermore, the differences across the methods are small.

A2. Technical Appendix

Lemma 1. *Suppose that Assumptions* 1 *and* 2 *hold. Define a profile estimator of* α_s *by*

$$\widehat{\alpha}_{s,T}(\alpha_w) = \underset{\alpha_s \in \mathbf{A}_s}{\operatorname{argmin}} \big(\widehat{\gamma}_T - f(\alpha_s, \alpha_w)\big)' \widehat{W}_T \big(\widehat{\gamma}_T - f(\alpha_s, \alpha_w)\big) \tag{A13}$$

TABLE A5. Frequencies of selecting model A (model B is misspecified and all impulse responses are used).

					QML on the RWMH Algorithm					Modified QML Based on the RWMH Algorithm				
					Geweke		SWZ			Geweke		SWZ		
T	H	Weight Matrix	$\widehat{q}_T$	QML Laplace	$\tau=0.5$	$\tau=0.9$	$q=0.5$	$q=0.9$	CJ	$\tau=0.5$	$\tau=0.9$	$q=0.5$	$q=0.9$	CJ
50	2	Diagonal	0.76	1.00	0.99	0.99	1.00	1.00	0.97	0.68	0.66	0.98	0.98	0.73
		Optimal	0.70	0.84	0.84	0.84	0.87	0.86	0.83	0.31	0.30	0.68	0.65	0.33
50	4	Diagonal	0.80	1.00	0.99	0.99	0.99	1.00	0.98	0.65	0.64	0.95	0.94	0.71
		Optimal	0.69	0.72	0.77	0.77	0.77	0.76	0.77	0.30	0.29	0.58	0.56	0.28
50	8	Diagonal	0.79	0.97	0.97	0.97	0.98	0.98	0.96	0.63	0.61	0.87	0.84	0.65
		Optimal	0.52	0.26	0.46	0.45	0.36	0.34	0.41	0.29	0.28	0.28	0.26	0.25
100	2	Diagonal	0.74	1.00	1.00	1.00	1.00	1.00	0.98	0.88	0.87	1.00	1.00	0.91
		Optimal	0.66	0.91	0.93	0.93	0.92	0.92	0.90	0.43	0.42	0.81	0.79	0.44
100	4	Diagonal	0.76	1.00	1.00	1.00	1.00	1.00	0.99	0.86	0.84	0.99	1.00	0.91
		Optimal	0.74	0.88	0.93	0.93	0.90	0.90	0.88	0.52	0.50	0.84	0.83	0.55
100	8	Diagonal	0.82	1.00	1.00	1.00	1.00	1.00	0.99	0.93	0.92	1.00	1.00	0.95
		Optimal	0.65	0.78	0.85	0.85	0.81	0.81	0.81	0.51	0.48	0.73	0.72	0.50
200	2	Diagonal	0.71	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Optimal	0.63	0.93	0.94	0.94	0.93	0.93	0.91	0.49	0.45	0.86	0.85	0.56
200	4	Diagonal	0.79	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Optimal	0.75	0.94	0.96	0.96	0.95	0.94	0.93	0.66	0.65	0.90	0.90	0.73
200	8	Diagonal	0.84	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	1.00	1.00	1.00
		Optimal	0.71	0.92	0.95	0.95	0.93	0.93	0.93	0.75	0.72	0.89	0.89	0.78

Note: T denotes the sample size and H denotes the maximum horizon for 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. "Optimal" refers to cases in which the weighting matrix is set to the inverse of the bootstrap covariance matrix of impulse responses. $\widehat{q}_T$ refers to the method that chooses the model whose estimation criterion function is smaller. "Laplace," "Geweke," "SWZ" and "CJ" refer to Laplace approximation, Geweke's (1999) modified harmonic mean estimator, Sims, Waggoner and Zha's (2008) modified harmonic mean estimator and the estimator of Chib and Jeliazkov (2001), respectively. τ and q are the tuning parameters for Geweke's (1999) and Sims, Waggoner, and Zha's (2008) methods. The numbers in the table are the frequencies of selecting Model A over Model B over 1000 Monte Carlo iterations.

TABLE A6. Frequencies of selecting model A (model A is more parsimonious and all impulse responses are used).

					QML on the RWMH Algorithm					Modified QML Based on the RWMH Algorithm				
					Geweke		SWZ			Geweke		SWZ		
T	H	Weight Matrix	$\widehat{q}_T$	QML Laplace	$\tau=0.5$	$\tau=0.9$	$q=0.5$	$q=0.9$	CJ	$\tau=0.5$	$\tau=0.9$	$q=0.5$	$q=0.9$	CJ
50	2	Diagonal	0.03	1.00	0.98	0.98	0.97	0.96	0.98	1.00	1.00	1.00	1.00	1.00
		Optimal	0.23	0.92	0.90	0.89	0.89	0.88	0.88	0.96	0.96	0.92	0.92	0.96
50	4	Diagonal	0.03	1.00	0.97	0.97	0.97	0.95	0.96	1.00	1.00	1.00	1.00	1.00
		Optimal	0.32	0.91	0.86	0.86	0.85	0.85	0.85	0.92	0.92	0.90	0.90	0.93
50	8	Diagonal	0.07	0.99	0.95	0.96	0.95	0.95	0.92	0.98	0.98	0.98	0.98	0.98
		Optimal	0.39	0.78	0.69	0.69	0.68	0.68	0.67	0.80	0.80	0.80	0.81	0.81
100	2	Diagonal	0.03	1.00	0.99	0.98	0.97	0.97	0.97	1.00	1.00	1.00	1.00	1.00
		Optimal	0.29	0.92	0.90	0.89	0.89	0.89	0.89	0.95	0.94	0.93	0.93	0.95
100	4	Diagonal	0.05	1.00	0.99	0.98	0.97	0.97	0.96	1.00	1.00	1.00	1.00	1.00
		Optimal	0.38	0.93	0.90	0.90	0.91	0.90	0.90	0.94	0.93	0.93	0.93	0.94
100	8	Diagonal	0.05	1.00	0.98	0.97	0.98	0.96	0.98	1.00	1.00	1.00	1.00	1.00
		Optimal	0.37	0.83	0.81	0.82	0.82	0.81	0.81	0.89	0.89	0.86	0.86	0.88
200	2	Diagonal	0.04	1.00	0.98	0.97	0.96	0.97	0.94	1.00	1.00	1.00	1.00	1.00
		Optimal	0.25	0.90	0.86	0.86	0.89	0.87	0.89	0.93	0.93	0.92	0.92	0.93
200	4	Diagonal	0.03	1.00	0.98	0.98	0.98	0.98	0.97	1.00	1.00	1.00	1.00	1.00
		Optimal	0.35	0.94	0.92	0.92	0.93	0.92	0.92	0.95	0.95	0.94	0.94	0.94
200	8	Diagonal	0.04	1.00	0.98	0.97	0.96	0.97	0.95	1.00	1.00	1.00	1.00	1.00
		Optimal	0.36	0.88	0.87	0.87	0.87	0.86	0.88	0.91	0.92	0.89	0.89	0.90

Note: See the notes for Table A5.

TABLE A7. Frequencies of selecting model A (model B is misspecified and only impulse responses to the technology shock are used).

					QML on the RWMH Algorithm					Modified QML Based on the RWMH Algorithm				
					Geweke		SWZ			Geweke		SWZ		
T	H	Weight Matrix	$\widehat{q}_T$	QML Laplace	$\tau=0.5$	$\tau=0.9$	$q=0.5$	$q=0.9$	CJ	$\tau=0.5$	$\tau=0.9$	$q=0.5$	$q=0.9$	CJ
50	2	Diagonal	1.00	1.00	0.98	0.98	0.99	0.99	0.97	0.60	0.58	0.96	0.95	0.62
		Optimal	0.98	0.99	0.97	0.97	1.00	1.00	0.97	0.41	0.38	0.87	0.83	0.45
50	4	Diagonal	0.98	1.00	0.99	0.99	1.00	1.00	0.98	0.65	0.63	0.97	0.97	0.68
		Optimal	0.92	0.93	0.93	0.93	0.94	0.93	0.92	0.37	0.36	0.83	0.81	0.41
50	8	Diagonal	0.95	0.99	0.99	0.99	0.97	0.98	0.98	0.63	0.63	0.93	0.92	0.64
		Optimal	0.64	0.46	0.61	0.61	0.54	0.53	0.56	0.31	0.30	0.49	0.47	0.29
100	2	Diagonal	1.00	1.00	1.00	1.00	1.00	1.00	0.97	0.78	0.77	1.00	1.00	0.84
		Optimal	0.99	1.00	1.00	1.00	1.00	1.00	0.98	0.67	0.65	0.99	0.98	0.75
100	4	Diagonal	1.00	1.00	1.00	1.00	1.00	1.00	0.97	0.83	0.81	1.00	1.00	0.87
		Optimal	0.97	0.99	1.00	1.00	1.00	1.00	0.98	0.70	0.67	0.98	0.98	0.77
100	8	Diagonal	0.99	1.00	1.00	1.00	1.00	1.00	0.98	0.84	0.83	1.00	1.00	0.89
		Optimal	0.84	0.85	0.91	0.91	0.86	0.86	0.89	0.68	0.66	0.81	0.81	0.68
200	2	Diagonal	1.00	1.00	1.00	1.00	1.00	1.00	0.97	0.92	0.91	1.00	1.00	0.94
		Optimal	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.89	0.87	1.00	1.00	0.94
200	4	Diagonal	0.99	1.00	1.00	1.00	1.00	1.00	0.98	0.91	0.91	1.00	1.00	0.96
		Optimal	0.97	1.00	1.00	1.00	1.00	1.00	0.98	0.91	0.90	1.00	1.00	0.92
200	8	Diagonal	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.94	0.93	0.99	1.00	0.96
		Optimal	0.85	0.96	0.98	0.98	0.96	0.96	0.96	0.84	0.83	0.90	0.89	0.86

Note: See the notes for Table A5.

TABLE A8. Frequencies of selecting model A (model A is more parsimonious and only impulse responses to the technology shock are used).

					QML on the RWMH Algorithm					Modified QML Based on the RWMH Algorithm				
					Geweke		SWZ			Geweke		SWZ		
T	H	Weight Matrix	$\widehat{q}_T$	QML Laplace	$\tau=0.5$	$\tau=0.9$	$q=0.5$	$q=0.9$	CJ	$\tau=0.5$	$\tau=0.9$	$q=0.5$	$q=0.9$	CJ
50	2	Diagonal	0.05	1.00	0.97	0.97	0.97	0.95	0.97	1.00	1.00	1.00	1.00	1.00
		Optimal	0.09	1.00	0.97	0.97	0.97	0.95	0.98	1.00	1.00	1.00	1.00	1.00
50	4	Diagonal	0.04	0.99	0.95	0.95	0.95	0.94	0.94	1.00	1.00	1.00	1.00	1.00
		Optimal	0.31	0.99	0.97	0.97	0.97	0.97	0.96	0.99	0.99	0.99	0.99	0.99
50	8	Diagonal	0.06	0.99	0.97	0.97	0.96	0.96	0.96	0.99	0.99	0.99	0.99	1.00
		Optimal	0.54	0.86	0.83	0.84	0.82	0.81	0.82	0.87	0.88	0.86	0.85	0.90
100	2	Diagonal	0.05	0.99	0.97	0.97	0.96	0.95	0.97	1.00	1.00	1.00	1.00	1.00
		Optimal	0.07	1.00	0.99	0.99	0.98	0.98	0.97	1.00	1.00	1.00	1.00	1.00
100	4	Diagonal	0.05	0.98	0.94	0.95	0.93	0.93	0.94	1.00	1.00	1.00	1.00	1.00
		Optimal	0.22	1.00	0.99	0.99	0.98	0.99	0.97	1.00	1.00	1.00	1.00	1.00
100	8	Diagonal	0.05	0.98	0.96	0.96	0.95	0.94	0.96	1.00	1.00	1.00	1.00	1.00
		Optimal	0.59	0.97	0.95	0.96	0.95	0.96	0.95	0.97	0.97	0.98	0.98	0.97
200	2	Diagonal	0.05	0.96	0.94	0.93	0.93	0.91	0.91	1.00	1.00	1.00	1.00	1.00
		Optimal	0.12	1.00	0.99	0.98	0.97	0.98	0.98	1.00	1.00	1.00	1.00	1.00
200	4	Diagonal	0.09	0.96	0.94	0.94	0.94	0.93	0.93	1.00	1.00	1.00	1.00	1.00
		Optimal	0.18	1.00	0.99	0.99	0.98	0.99	0.99	1.00	1.00	1.00	1.00	1.00
200	8	Diagonal	0.07	0.98	0.97	0.96	0.95	0.95	0.94	1.00	1.00	1.00	1.00	1.00
		Optimal	0.63	0.98	0.97	0.97	0.97	0.97	0.97	0.99	0.99	0.98	0.99	0.98

Note: See the notes for Table A5.

TABLE A9. The mean and standard deviation of QML estimates.

		Geweke				SWZ					
T	H	$\tau=0.5$		$\tau=0.9$		$q=0.5$		$q=0.9$		CJ	
		All impulse responses are used (Case 1, Model A)									
50	2	−44.51	[0.75]	−44.27	[0.73]	−41.83	[0.68]	−42.29	[0.76]	−41.60	[0.68]
50	4	−69.38	[7.91]	−68.89	[7.93]	−72.32	[7.39]	−72.17	[7.26]	−67.06	[7.36]
50	8	−148.54	[5.00]	−148.03	[5.07]	−151.33	[2.12]	−151.27	[2.13]	−146.40	[4.32]
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 (Case 3, Model A)									
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]

Note: The means and standard deviations (in brackets) in each row are calculated from 100 QML estimates given a realization of data. See the notes to Table A5.

for each $\alpha_w \in \mathbf{A}_w$. *Then*

$$\sup_{\alpha_w \in \mathbf{A}_w} \left\| \widehat{\alpha}_s(\alpha_w) - \alpha_{s,0} \right\| = O_p\left(T^{-\frac{1}{2}}\right), \quad \text{(A14)}$$

$$\sup_{\alpha_w \in \mathbf{A}_w} \left| \widehat{q}_{A,T}\left(\widehat{\alpha}_{s,T}(\alpha_w), \alpha_w\right) - \widehat{q}_{A,T}\left(\widehat{\alpha}_{s,T}(\alpha_{w,0}), \alpha_{w,0}\right) \right| = O_p\left(T^{-\frac{1}{2}}\right), \quad \text{(A15)}$$

$$\sup_{\alpha_w \in \mathbf{A}_w} \left\| \operatorname{vech}\left(\nabla^2_{\alpha_s} \widehat{q}_{A,T}\left(\widehat{\alpha}_{s,T}(\alpha_w), \alpha_w\right) - \nabla^2_{\alpha_s} \widehat{q}_{A,T}\left(\widehat{\alpha}_{s,T}(\alpha_{w,0}), \alpha_{w,0}\right)\right) \right\| = O_p\left(T^{-\frac{1}{2}}\right). \quad \text{(A16)}$$

PROOF OF LEMMA 1. The pointwise convergence of $\widehat{\alpha}_s(\alpha_w)$ to α_s,

$$\widehat{\alpha}_{s,T}(\alpha_w) \xrightarrow{p} \alpha_{s,0} \quad \text{(A17)}$$

for each $\alpha_w \in \mathbf{A}_w$, follows from Assumption 1(a), (b), and 2(b). $\widehat{\alpha}_{s,T}(\alpha_w)$ satisfies the first-order conditions:

$$\nabla_{\alpha_s} q_A\left(\widehat{\alpha}_{s,T}(\alpha_w), \alpha_w\right) = 0_{p_{A_s} \times 1}. \quad \text{(A18)}$$

Let $\widehat{J}_s$ and $\widehat{J}_w$ denote the Jacobian matrices of the left-hand side of (A18) with respect to α_s and α_w, respectively. It follows from the pointwise convergence, Assumptions 1(b), 2(a), and 2(c), $\widehat{J}_s$ is nonsingular with probability approaching one. Thus, applying the implicit function theorem to (A18) yields

$$\frac{\partial \widehat{\alpha}_{s,T}(\alpha_w)}{\partial \alpha_w'} = -\widehat{J}_s^{-1}\widehat{J}_w = O_p\big(T^{-\frac{1}{2}}\big), \tag{A19}$$

where $O_p(T^{-1/2})$ is uniform in $\alpha_w \in \mathbf{A}_w$ which follows from Assumptions 1(a), (b) and 2(a). It follows from the mean value theorem and (A19) that

$$\widehat{\alpha}_{s,T}\big(\alpha_w'\big) - \widehat{\alpha}_{s,T}(\alpha_w) = O_p\big(T^{-\frac{1}{2}}\left\|\alpha_w' - \alpha_w\right\|\big). \tag{A20}$$

Given the pointwise convergence, the compactness of $\mathbf{A}_w$ and stochastic equicontinuity (A20), we can strengthen the pointwise convergence to uniform convergence (A14) by Theorem 1 of Andrews (1992). Then (A15) and (A16) follow from (A14), Assumptions 1(b) and 2(a). □

Proof of Theorem 1. Let

$$B_\epsilon(\widehat{\alpha}_{s,T}) = \big\{\alpha_s \in \mathbf{A}_s : \|\alpha_s - \widehat{\alpha}_{s,T}\| < \epsilon\big\},$$

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

$$m_A = \int_{B_\epsilon(\widehat{\alpha}_{s,T})\times A_w} \pi_A(\alpha)e^{-T\widehat{q}_{A,T}(\alpha)}\, d\alpha + \int_{(\mathbf{A}_s\setminus B_\epsilon(\widehat{\alpha}_{s,T}))\times A_w} \pi_A(\alpha)e^{-T\widehat{q}_{A,T}(\alpha)}\, d\alpha. \tag{A21}$$

It follows from Taylor's theorem, the first-order condition for $\widehat{\alpha}_{s,T}(\alpha_w)$, (A14) and (A16) that

$$\begin{aligned}
\widehat{q}_{A,T}(\alpha) &= \widehat{q}_{A,T}\big(\widehat{\alpha}_{s,T}(\alpha_w), \alpha_w\big) + \nabla_{\alpha_s}\widehat{q}_{A,T}\big(\widehat{\alpha}_{s,T}(\alpha_w), \alpha_w\big)'\big(\alpha_s - \widehat{\alpha}_{s,T}(\alpha_w)\big)\\
&\quad + \frac{1}{2}\big(\alpha_s - \widehat{\alpha}_{s,T}(\alpha_w)\big)'\nabla^2_{\alpha_s}\widehat{q}_{A,T}\big(\bar{\alpha}_{s,T}(\alpha), \alpha_w\big)\big(\alpha_s - \widehat{\alpha}_{s,T}(\alpha_w)\big)\\
&= \widehat{q}_{A,T}\big(\widehat{\alpha}_{s,T}(\alpha_w), \alpha_w\big)\\
&\quad + \frac{1}{2}\big(\alpha_s - \widehat{\alpha}_{s,T}(\alpha_w)\big)'\nabla^2_{\alpha_s}\widehat{q}_{A,T}\big(\bar{\alpha}_{s,T}(\alpha_w), \alpha_w\big)\big(\alpha_s - \widehat{\alpha}_{s,T}(\alpha_w)\big)\\
&= \widehat{q}_{A,T}(\widehat{\alpha}_T) + \frac{1}{2}(\alpha_s - \widehat{\alpha}_{s,T})'\nabla^2_{\alpha_s}\widehat{q}_{A,T}(\widehat{\alpha}_T)(\alpha_s - \widehat{\alpha}_{s,T}) + O_p\big(T^{-\frac{1}{2}}\big),
\end{aligned} \tag{A22}$$

uniformly in $\alpha_w \in \mathbf{A}_w$, where $\bar{\alpha}_{s,T}(\alpha_w)$ is a point between α_s and $\widehat{\alpha}_{s,T}(\alpha_w)$.

Using (A22), Lemma 1 and Assumption 1(c) the first integral on the right-hand side of (A21) can be written as:

$$\begin{aligned}
&\int_{B_\epsilon(\widehat{\alpha}_{s,T})\times\mathbf{A}_w} \pi_A(\alpha)\\
&\qquad \times e^{-T\widehat{q}_{A,T}(\widehat{\alpha}_{s,T}(\alpha_w),\alpha_w) - \frac{T}{2}(\alpha_s - \widehat{\alpha}_{s,T}(\alpha_w))'\nabla^2_{\alpha_s}\widehat{q}_{A,T}(\bar{\alpha}_{s,T}(\alpha),\alpha_w)(\alpha_s - \widehat{\alpha}_{s,T}(\alpha_w))}\, d\alpha\big(1 + o_p(1)\big)\\
&= e^{-T\widehat{q}_{A,T}(\widehat{\alpha}_T)}\int_{B_\epsilon(\widehat{\alpha}_{s,T})\times\mathbf{A}_w} \pi_A(\alpha)e^{-\frac{T}{2}(\alpha_s - \widehat{\alpha}_{s,T})'\nabla^2_{\alpha_s}\widehat{q}_{A,T}(\widehat{\alpha}_T)(\alpha_s - \widehat{\alpha}_{s,T})}\, d\alpha\big(1 + o_p(1)\big)
\end{aligned}$$

$$
\begin{aligned}
&= e^{-T\widehat{q}_{A,T}(\widehat{\alpha}_T)} \int_{B_\epsilon(\widehat{\alpha}_{s,T})\times \mathbf{A}_w} \pi_A(\widehat{\alpha}_{s,T}, \alpha_w) \\
&\quad \times e^{-\frac{T}{2}(\alpha_s-\widehat{\alpha}_{s,T})'\nabla^2_{\alpha_s}\widehat{q}_{A,T}(\widehat{\alpha}_T)(\alpha_s-\widehat{\alpha}_{s,T})}\, d\alpha \big(1+o_p\big(e^{-\frac{T}{2}\lambda_T\epsilon^2}\big)\big) \qquad \text{(A23)}\\
&= \pi_{A_s}(\widehat{\alpha}_{s,T}) e^{-T\widehat{q}_{A,T}(\widehat{\alpha}_T)} \left[\int_{B_\epsilon(\widehat{\alpha}_{s,T})} e^{-\frac{T}{2}(\alpha_s-\widehat{\alpha}_{s,T})'\nabla^2_{\alpha_s}\widehat{q}_{A,T}(\widehat{\alpha}_T)(\alpha_s-\widehat{\alpha}_{s,T})}\, d\alpha_s\right] \\
&\quad \times \left[\int_{A_w} \pi_{A_w|A_s}(\alpha_w|\widehat{\alpha}_{s,T})\, d\alpha_w\right] \big(1+o_p(1)\big) \\
&= \pi_{A_s}(\widehat{\alpha}_T) e^{-T\widehat{q}_{A,T}(\widehat{\alpha}_{s,T})} \left(\frac{2\pi}{T}\right)^{\frac{p_{A_s}}{2}} \left|\nabla^2_{\alpha_s}\widehat{q}_{A,T}(\widehat{\alpha}_T)\right|^{-\frac{1}{2}} \\
&\quad \times \big(1+o_p(1)\big),
\end{aligned}
$$

where λ_T is a sequence of strictly positive bounded constants and $o_p(e^{-\frac{T}{2}\lambda_T\epsilon^2})$ is uniform on $B_\epsilon(\widehat{\alpha}_T)$ and $\pi_{A_w|A_s}(\alpha_w|\alpha_s)$ is the prior of α_w conditional on α_s. Thus, the first integral on the right-hand side of (A21) can be approximated by

$$
\pi_{A_s}(\widehat{\alpha}_{s,T}) e^{-T\widehat{q}_{A,T}(\widehat{\alpha}_{s,T})} \left(\frac{2\pi}{T}\right)^{\frac{p_{A_s}}{2}} \left|\nabla^2_{\alpha_s}\widehat{q}_{A,T}(\widehat{\alpha}_T)\right|^{-\frac{1}{2}}. \qquad \text{(A24)}
$$

By letting $\epsilon \to 0$, it follows from Assumptions 1(b) and 2(b) that the second integral on the right-hand side of (A21) can be bounded as

$$
\begin{aligned}
&\left|\int_{(A_s\setminus B_\epsilon(\widehat{\alpha}_T))\times \mathbf{A}_w} \pi_A(\alpha) e^{-T\widehat{q}_{A,T}(\alpha)}\, d\alpha\right| \\
&\quad \le \int_{(\mathbf{A}_s\setminus B_\epsilon(\widehat{\alpha}_T))\times \mathbf{A}_w} \pi_A(\alpha)\, d\alpha \times e^{-T \inf_{\alpha\in(\mathbf{A}_s\setminus B_\epsilon(\widehat{\alpha}_T))\times \mathbf{A}_w} \widehat{q}_{A,T}(\alpha)} \qquad \text{(A25)}\\
&\quad = O\big(e^{-T(\widehat{q}_{A,T}(\widehat{\alpha}_T)+\eta)}\big),
\end{aligned}
$$

for some $\eta > 0$. Combining (A24) and (A25), we obtain

$$
m_A = \pi_{A_s}(\widehat{\alpha}_{s,T}) e^{-T\widehat{q}_{A,T}(\widehat{\alpha}_T)} \left(\frac{2\pi}{T}\right)^{\frac{p_{A_s}}{2}} \left|\nabla^2_{\alpha_s}\widehat{q}_{A,T}(\widehat{\alpha}_T)\right|^{-\frac{1}{2}} \big(1+o_p(1)\big). \qquad \text{(A26)}
$$

When there is no strongly identified parameter ($p_{A_s} = 0$), it follows from Assumption 2(a) that

$$
m_A = O_p(1). \qquad \text{(A27)}
$$

Similarly,

$$
m_B = \begin{cases} \pi_{B_s}(\widehat{\beta}_{s,T}) e^{-T\widehat{q}_{B,T}(\widehat{\beta}_T)} \left(\frac{2\pi}{T}\right)^{\frac{p_{B_s}}{2}} \left|\nabla^2_{\beta_s}\widehat{q}_{B,T}(\widehat{\beta}_T)\right|^{-\frac{1}{2}} \big(1+o_p(1)\big) & \text{if } p_{B_s} > 0, \\ O_p(1) & \text{if } p_{B_s} = 0. \end{cases} \qquad \text{(A28)}
$$

Theorem 1 follows from (A26)–(A28) and Assumptions 1(b)(c) and 2(c). □

We will use the following lemma in the proof of Theorem 2:

LEMMA 2. *Suppose that Assumptions* 1 *and* 3 *hold. Define a profile estimator of* α_s *by*

$$\widehat{\alpha}_{s,T}(\alpha_p) = \underset{\alpha_s \in \mathbf{A}_s}{\operatorname{argmin}} \big(\widehat{\gamma}_T - f(\alpha_s, \alpha_p)\big)' \widehat{W}_T \big(\widehat{\gamma}_T - f(\alpha_s, \alpha_p)\big), \tag{A29}$$

for each $\alpha_p \in \mathbf{A}_p$. *Then*

$$\sup_{\alpha_p \in \mathbf{A}_{p,0}} \big\| \widehat{\alpha}_s(\alpha_p) - \alpha_{s,0} \big\| = o_p(1), \tag{A30}$$

$$\sup_{\alpha_p \in \mathbf{A}_{p,0}} \big| \widehat{q}_{A,T}\big(\widehat{\alpha}_{s,T}(\alpha_p), \alpha_p\big) - \widehat{q}_{A,T}\big(\widehat{\alpha}_{s,T}(\alpha_{p,0}), \alpha_{p,0}\big) \big| = o_p(1), \tag{A31}$$

$$\sup_{\alpha_p \in \mathbf{A}_{p,0}} \big\| \operatorname{vech}\big(\nabla^2_{\alpha_s} \widehat{q}_{A,T}\big(\widehat{\alpha}_{s,T}(\alpha_p), \alpha_p\big) - \nabla^2_{\alpha_s} \widehat{q}_{A,T}\big(\widehat{\alpha}_{s,T}(\alpha_{p,0}), \alpha_{p,0}\big)\big) \big\| = O_p\big(T^{-\frac{1}{2}}\big). \tag{A32}$$

PROOF OF LEMMA 2. The pointwise convergence of $\widehat{\alpha}_s(\alpha_p)$ follows from the usual arguments. Note that $\widehat{\alpha}_{s,T}(\alpha_p)$ satisfies the first-order conditions:

$$F_{\alpha_s}\big(\widehat{\alpha}_{s,T}(\alpha_p), \alpha_p\big)' \widehat{W}_T \big(\widehat{\gamma}_T - f\big(\widehat{\alpha}_{s,T}(\alpha_p), \alpha_p\big)\big) = 0_{p_{A_s} \times 1}, \tag{A33}$$

for every $\alpha_p \in \mathbf{A}_{p,0}$. Let $\widehat{J}_s$ and $\widehat{J}_p$ denote the Jacobian matrices of the left-hand side of (A33) with respect to α_s and α_p, respectively. By Assumptions 1(b) and 3(c), $\widehat{J}_s$ is nonsingular with probability approaching one. Thus, it follows from the implicit function theorem that

$$\frac{\partial \widehat{\alpha}_{s,T}(\alpha_p)}{\partial \alpha_p'} = -\widehat{J}_s^{-1} \widehat{J}_p = O_p(1). \tag{A34}$$

It follows from the mean value theorem and (A34) that

$$\widehat{\alpha}_{s,T}\big(\alpha_p'\big) - \widehat{\alpha}_{s,T}(\alpha_p) = O_p\big(\big\|\alpha_p' - \alpha_p\big\|\big), \tag{A35}$$

where $\alpha_p, \alpha_p' \in \mathbf{A}_{p,0}$. Given the pointwise convergence $\widehat{\alpha}_{s,T}(\alpha_p) \xrightarrow{p} \alpha_{s,0}$ for each $\alpha_p \in \mathbf{A}_{p,0}$, the compactness of $\mathbf{A}_p$ and stochastic equicontinuity (A35), we can strengthen the pointwise convergence to uniform convergence (A30) by Theorem 1 of Andrews (1992).

It follows from Assumption 1(b), Lemma 2(a), and the definition of $\mathbf{A}_{p,0}$ that

$$\begin{aligned}
&\big| \widehat{q}_{A,T}\big(\widehat{\alpha}_{s,T}(\alpha_p), \alpha_p\big) - \widehat{q}_{A,T}\big(\widehat{\alpha}_{s,T}(\alpha_{p,0}), \alpha_{p,0}\big) \big| \\
&\le \big| \widehat{q}_{A,T}\big(\widehat{\alpha}_{s,T}(\alpha_p), \alpha_p\big) - q_A\big(\widehat{\alpha}_{s,T}(\alpha_p), \alpha_p\big) \big| \\
&\quad + \big| \widehat{q}_{A,T}\big(\widehat{\alpha}_{s,T}(\alpha_{p,0}), \alpha_{p,0}\big) - q_A\big(\widehat{\alpha}_{s,T}(\alpha_{p,0}), \alpha_{p,0}\big) \big| \\
&\quad + \big| q_A\big(\widehat{\alpha}_{s,T}(\alpha_p), \alpha_p\big) - q_A\big(\widehat{\alpha}_{s,T}(\alpha_{p,0}), \alpha_{p,0}\big) \big| \\
&= o_p(1)
\end{aligned}$$

uniformly in $\alpha_p \in \mathbf{A}_{p,0}$, from which (A31) follows. (A32) follows from similar arguments. □

PROOF OF THEOREM 2(A). We can write

$$
\begin{aligned}
&\int_{\mathbf{A}} \pi_A(\alpha)\exp\big(-T\widehat{q}_{A,T}(\alpha)\big)\,d\alpha \\
&\quad = \int_{\mathbf{A}_0} \pi_A(\alpha)\exp\big(-T\widehat{q}_{A,T}(\alpha)\big)\,d\alpha \\
&\qquad + \int_{(\mathbf{A}_0^c)^{-\varepsilon}} \pi_A(\alpha)\exp\big(-T\widehat{q}_{A,T}(\alpha)\big)\,d\alpha \\
&\qquad + \int_{\mathbf{A}\setminus(\mathbf{A}_0\cup(\mathbf{A}_0^c)^{-\varepsilon})} \pi_A(\alpha)\exp\big(-T\widehat{q}_{A,T}(\alpha)\big)\,d\alpha \\
&\quad = I_1 + I_2 + I_3, \quad \text{say.}
\end{aligned} \tag{A36}
$$

It follows from Assumption 1(b) that

$$
\begin{aligned}
I_1 &= \int_{\mathbf{A}_0} \pi_A(\alpha)\exp\big(-Tq_A(\alpha)\big)\,d\alpha\big(1+o_p(1)\big) \\
&= \int_{\mathbf{A}_0} \pi_A(\alpha)\,d\alpha\exp\big(-Tq_A(\alpha_0)\big)\big(1+o_p(1)\big),
\end{aligned} \tag{A37}
$$

for any $\alpha_0 \in \mathbf{A}_0$. It follows from Assumptions 1(b) and 3(a) that

$$
I_2 = o_p\big(\exp\big(-Tq_A(\alpha_0)\big)\big). \tag{A38}
$$

By letting $\varepsilon \to 0$, the term I_3 can be made arbitrarily small. Combining (A36)–(A38), we can approximate the QML for model A by

$$
m_A = \int_{\mathbf{A}_0} \pi_A(\alpha)\,d\alpha \times \exp\big(-Tq_A(\alpha_0)\big)\big(1+o_p(1)\big) \tag{A39}
$$

for any $\alpha_0 \in \mathbf{A}_0$. Similarly, the QML for model B can be approximated by

$$
m_B = \int_{\mathbf{B}_0} \pi_B(\beta)\,d\beta \times \exp\big(-Tq_B(\beta_0)\big)\big(1+o_p(1)\big), \tag{A40}
$$

for any $\beta_0 \in \mathbf{B}_0$. Theorem 2(a) follows from (A39) and (A40). □

PROOF OF THEOREM 2(B). Define

$$
B_\epsilon(\widehat{\alpha}_{s,T}) = \big\{\alpha_s \in \mathbf{A}_s : \|\alpha_s - \widehat{\alpha}_{s,T}\| < \epsilon\big\}.
$$

The QML of model A can be written as

$$
\begin{aligned}
&\int_{\mathbf{A}} \pi_A(\alpha)\exp\big(-T\widehat{q}_{A,T}(\alpha)\big)\,d\alpha \\
&\quad = \int_{B_\epsilon(\widehat{\alpha}_{s,T})\times\mathbf{A}_{p,0}} \pi_A(\alpha)\exp\big(-T\widehat{q}_{A,T}(\alpha)\big)\,d\alpha
\end{aligned}
$$

$$+ \int_{((B_\epsilon(\widehat{\alpha}_{s,T}) \times \mathbf{A}_{p,0})^c)^{-\varepsilon}} \pi_A(\alpha) \exp\big(-T\widehat{q}_{A,T}(\alpha)\big)\, d\alpha \tag{A41}$$

$$+ \int_{\mathbf{A} \setminus ((B_\epsilon(\widehat{\alpha}_{s,T}) \times \mathbf{A}_0) \cup (B_\epsilon(\widehat{\alpha}_{s,T}) \times \mathbf{A}_0)^c))^{-\varepsilon})} \pi_A(\alpha) \exp\big(-T\widehat{q}_{A,T}(\alpha)\big)\, d\alpha$$

$$= I_1 + I_2 + I_3, \quad \text{say.}$$

It follows from Assumption 1(b) and Lemmas 2(a), (b) that

$$\begin{aligned}
I_1 &= \int_{B_\epsilon(\widehat{\alpha}_{s,T}) \times \mathbf{A}_{p,0}} \pi_A(\alpha) e^{-T\widehat{q}_{A,T}(\alpha)}\, d\alpha \\
&= \int_{B_\epsilon(\widehat{\alpha}_{s,T}) \times \mathbf{A}_{p,0}} \pi_A(\alpha) e^{-T\widehat{q}_{A,T}([\alpha_s' \alpha'_{p,0}]')}\, d\alpha (1+\epsilon) \\
&= \int_{B_\epsilon(\widehat{\alpha}_{s,T}) \times \mathbf{A}_{p,0}} \pi_A(\alpha) e^{-T\widehat{q}_{A,T}(\widehat{\alpha}_{s,T}(\alpha_{p,0}), \alpha_{p,0})} \\
&\quad \times e^{-\frac{T}{2}(\alpha_s - \widehat{\alpha}_{s,T}(\alpha_{p,0}))' \nabla^2_{\alpha_s} \widehat{q}_{A,T}(\widehat{\alpha}_{s,T}(\alpha_{p,0}), \alpha_{p,0})(\alpha_s - \widehat{\alpha}_{s,T}(\alpha_{p,0}))}\, d\alpha \\
&\quad \times \big(1 + o_p(\epsilon)\big) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(A42)} \\
&= e^{-T\widehat{q}_{A,T}([\widehat{\alpha}_T(\alpha_{p,0})' \alpha'_{p,0}]')} \left(\frac{2\pi}{T}\right)^{\frac{p_{A_s}}{2}} \big\| \nabla^2_{\alpha_s} \widehat{q}_{A,T}\big([\widehat{\alpha}_T(\alpha_{p,0})' \alpha'_{p,0}]'\big) \big\|^{-\frac{1}{2}} \\
&\quad \times \int_{\mathbf{A}_{p,0}} \pi_A(\alpha_{s,0}, \alpha_p)\, d\alpha_p \big(1 + o_p(\epsilon)\big) \\
&= e^{-T\widehat{q}_{A,T}(\widehat{\alpha}_T)} \left(\frac{2\pi}{T}\right)^{\frac{p_{A_s}}{2}} \big\| \nabla^2_{\alpha_s} \widehat{q}_{A,T}\big([\widehat{\alpha}_T(\alpha_{p,0})' \alpha'_{p,0}]'\big) \big\|^{-\frac{1}{2}} \\
&\quad \times \int_{\mathbf{A}_{p,0}} \pi_A(\alpha_{s,0}, \alpha_p)\, d\alpha_p \big(1 + o_p(\epsilon)\big).
\end{aligned}$$

Thus, the QML can be approximated by

$$\begin{aligned}
m_A &= e^{-T\widehat{q}_{A,T}(\widehat{\alpha}_T)} \left(\frac{2\pi}{T}\right)^{\frac{p_{A_s}}{2}} \big| \nabla^2_{\alpha_s} \widehat{q}_{A,T}\big([\widehat{\alpha}_{s,T}(\alpha_{p,0})' \alpha'_{p,0}]'\big) \big|^{-\frac{1}{2}} \\
&\quad \times \int_{\mathbf{A}_{p,0}} \pi_A(\alpha_{s,0}, \alpha_p)\, d\alpha_p \big(1 + o_p(1)\big),
\end{aligned} \tag{A43}$$

as $\epsilon \to 0$. The rest of the proof is analogous to that of Theorem 1. □

Proof of Proposition 1(a). Because

$$\nabla_{\alpha_s} \widehat{q}_{A,T}(\alpha) = 2F_s(\alpha_s)' \widehat{W}_T \big(\widehat{\gamma}_T - f(\alpha)\big), \tag{A44}$$

$$\nabla_{\alpha_s} q_A(\alpha) = 2F_s(\alpha_s)' W \big(\gamma - f(\alpha)\big), \tag{A45}$$

it follows from the compactness of $\mathbf{A}$, the twice continuous differentiability of f and $\widehat{W}_T \xrightarrow{p} W$ that

$$\begin{aligned}\nabla_{\alpha_s}\widehat{q}_{A,T}(\alpha) - \nabla_{\alpha_s} q_A(\alpha) &= -2F_s(\alpha_s)'\big[\widehat{W}_T\big(\widehat{\gamma}_T - f(\alpha)\big) - W\big(\gamma - f(\alpha)\big)\big] \\ &= -2F_s(\alpha_s)'\big[(\widehat{W}_T - W)\big(\gamma - f(\alpha)\big) + \widehat{W}_T(\widehat{\gamma}_T - \gamma)\big] \\ &= o_p(1),\end{aligned} \tag{A46}$$

uniformly in $\alpha_s \in \mathbf{A}_s$ as required in Assumption 1(b). Similarly, because

$$\nabla^2_{\alpha_s}\widehat{q}_{A,T}(\alpha) = 2F_s(\alpha_s)'\widehat{W}_T F_s(\alpha_s) - 2\big[\big(\widehat{\gamma}_T - f_s(\alpha_s)\big)'\widehat{W}_T \otimes I_{p_{A_s}}\big]\frac{\partial\,\mathrm{vec}\big(F_s(\alpha_s)'\big)}{\partial\alpha_s'}, \tag{A47}$$

$$\nabla^2_{\alpha_s} q_A(\alpha) = 2F_s(\alpha_s)'WF_s(\alpha_s) - 2\big[\big(\gamma - f_s(\alpha_s)\big)'W \otimes I_{p_{A_s}}\big]\frac{\partial\,\mathrm{vec}\big(F_s(\alpha_s)'\big)}{\partial\alpha_s'}, \tag{A48}$$

it follows that

$$\begin{aligned}&\nabla_{\alpha_s}\widehat{q}_A(\alpha) - \nabla_{\alpha_s} q_A(\alpha) \\ &\quad = 2F_s(\alpha_s)'(\widehat{W}_T - W)F_s(\alpha_s) \\ &\qquad - 2\big[\big(\gamma - f_s(\alpha_s)\big)'(\widehat{W}_T - W) \otimes I_{p_{A_s}}\big]\frac{\partial\,\mathrm{vec}\big(F_s(\alpha_s)'\big)}{\partial\alpha_s'} \\ &\qquad - 2\big[(\widehat{\gamma}_T - \gamma)'\widehat{W}_T \otimes I_{p_{A_s}}\big]\frac{\partial\,\mathrm{vec}\big(F_s(\alpha_s)'\big)}{\partial\alpha_s'} \\ &\quad = o_p(1),\end{aligned} \tag{A49}$$

uniformly in $\alpha_s \in \mathbf{A}_s$ as in Assumption 1(b). Because $F_s(\alpha_{s,0})$ has rank p_{A_s} and W is positive definite,

$$\begin{aligned}&\nabla^2_{\alpha_s} q_A(\alpha_0) \\ &\quad = 2F_s(\alpha_{s,0})'WF_s(\alpha_{s,0}) - 2\big[\big(\gamma - f_s(\alpha_{s,0})\big)'\widehat{W}_T \otimes I_{p_{A_s}}\big]\frac{\partial\,\mathrm{vec}\big(F_s(\alpha_{s,0})'\big)}{\partial\alpha_{s,0}'} \\ &\quad = 2F_s(\alpha_{s,0})'WF_s(\alpha_{s,0})\end{aligned} \tag{A50}$$

is nonsingular as required in Assumption 2(c). Because

$$\begin{aligned}\widehat{q}_{A,T}(\alpha) &= \big(\widehat{\gamma}_T - f_s(\alpha_s) - T^{-\frac{1}{2}}f_w(\alpha)\big)'\widehat{W}_T\big(\widehat{\gamma}_T - f_s(\alpha_s) - T^{-\frac{1}{2}}f_w(\alpha)\big) \\ &= (\widehat{\gamma}_T - f_s(\alpha_s)'\widehat{W}_T(\widehat{\gamma}_T - f_s(\alpha_s) + 2T^{-\frac{1}{2}}\big(\widehat{\gamma}_T - f_s(\alpha_s)\big)'\widehat{W}_T f_w(\alpha) \\ &\quad + T^{-1}f_w(\alpha)'\widehat{W}_T f_w(\alpha),\end{aligned} \tag{A51}$$

$q_A(\alpha)$ can be written as $q_{A_s}(\alpha_s) + T^{-1/2}q_{A_w}(\alpha)$ where $q_{A_s}(\alpha_s) = (\gamma - f_s(\alpha_s))'W(\gamma - f_s(\alpha_s))$ and $q_{A_w}(\alpha)$ is defined as

$$T^{\frac{1}{2}}\big(\widehat{\gamma}_T - f_s(\alpha_s)\big)'Wf_w(\alpha) + f_w(\alpha)'Wf_w(\alpha) \Rightarrow q_{A_w}(\alpha) \tag{A52}$$

and is $O_p(1)$ uniformly in $\alpha \in \mathbf{A}$ as in Assumption 2(a). Because W is positive definite, the uniqueness of α_s imply Assumption 2(b). □

Proof of Proposition 1(b). Because

$$\nabla_{\alpha_s}\widehat{q}_{A,T}(\alpha) = 2F_s(\alpha)'\widehat{W}_T\big(\widehat{\gamma}_T - f(\alpha)\big), \tag{A53}$$

$$\nabla_{\alpha_s}q_A(\alpha) = 2F_s(\alpha)'W\big(\gamma_T - f(\alpha)\big), \tag{A54}$$

it follows from the compactness of $\mathbf{A}$, the twice continuous differentiability of f and $\widehat{W}_T \xrightarrow{p} W$ that

$$\begin{aligned}\nabla_{\alpha_s}\widehat{q}_{A,T}(\alpha) - \nabla_{\alpha_s}q_A(\alpha) &= -2F_s(\alpha)'\big[\widehat{W}_T\big(\widehat{\gamma}_T - f(\alpha)\big) - W\big(\widehat{\gamma}_T - f(\alpha)\big)\big]\\ &= -2F_s(\alpha)'\big[(\widehat{W}_T - W)\big(\gamma - f(\alpha)\big) + \widehat{W}_T(\widehat{\gamma}_T - \gamma)\big] \\ &= o_p(1),\end{aligned} \tag{A55}$$

uniformly in $\alpha \in \mathbf{A}$ as required in Assumption 1(b). Similarly, because

$$\nabla^2_{\alpha_s}\widehat{q}_{A,T}(\alpha) = 2F_s(\alpha)'\widehat{W}_T F_s(\alpha) - 2\big[\big(\widehat{\gamma}_T - f(\alpha)\big)'\widehat{W}_T \otimes I_{p_{A_s}}\big]\frac{\partial \operatorname{vec}\big(F_s(\alpha)'\big)}{\partial \alpha_s'}, \tag{A56}$$

$$\nabla^2_{\alpha_s}q_A(\alpha) = 2F_s(\alpha)'WF_s(\alpha) - 2\big[\big(\gamma - f(\alpha)\big)'W \otimes I_{p_{A_s}}\big]\frac{\partial \operatorname{vec}\big(F_s(\alpha)'\big)}{\partial \alpha_s'}, \tag{A57}$$

it follows that

$$\begin{aligned}&\nabla^2_{\alpha_s}\widehat{q}_A(\alpha) - \nabla^2_{\alpha_s}q_A(\alpha)\\ &\quad= 2F_s(\alpha)'(\widehat{W}_T - W)F_s(\alpha)\\ &\qquad - 2\big[\big(\gamma - f(\alpha_s)\big)'(\widehat{W}_T - W) \otimes I_{p_{A_s}}\big]\frac{\partial \operatorname{vec}\big(F_s(\alpha)'\big)}{\partial \alpha_s'}\\ &\qquad - 2\big[(\widehat{\gamma}_T - \gamma)'\widehat{W}_T \otimes I_{p_{A_s}}\big]\frac{\partial \operatorname{vec}\big(F_s(\alpha)'\big)}{\partial \alpha_s'} \qquad\qquad (A58)\\ &\quad= o_p(1), \qquad\qquad (A59)\end{aligned}$$

uniformly in $\alpha_s \in \mathbf{A}_s$ as in Assumption 1(b). Assumption 4(h) implies that $\nabla^2_{\alpha_s}q_A(\alpha_0)$ is positive definite as required by Assumption 3(b). Because W is positive definite, the uniqueness of α_s imply Assumption 3(b). □

Proof of Proposition 2(a). Because

$$\nabla_{\alpha_s}\widehat{q}_{A,T}(\alpha) = 2\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial f(x_t,\alpha)}{\partial \alpha'}\right]'\widehat{W}_{A,T}\frac{1}{T}\sum_{t=1}^{T}f(x_t,\alpha), \tag{A60}$$

$$\nabla_{\alpha_s}q_A(\alpha) = 2F_s(\alpha_s)'W_A\big(f_s(\alpha_s)\big), \tag{A61}$$

and because $(1/T)\sum_{t=1}^{T}\partial f(x_t,\alpha)/\partial\alpha' - F_s(\alpha_s) = o_p(1)$ and $(1/T)\sum_{t=1}^{T} f(x_t,\alpha) - f_s(\alpha_s) = o_p(1)$ and $\widehat{W}_{A,T} \xrightarrow{p} W_A$, it follows from the compactness of **A** and the three times continuous differentiability of f that

$$\nabla_{\alpha_s}\widehat{q}_{A,T}(\alpha) - \nabla_{\alpha_s} q_A(\alpha) = o_p(1), \tag{A62}$$

uniformly in $\alpha_s \in \mathbf{A}_s$ as required in Assumption 1(b). Similarly, because

$$\begin{aligned}
&\nabla^2_{\alpha_s}\widehat{q}_{A,T}(\alpha)\\
&= 2\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial f(x_t,\alpha)}{\partial\alpha'}\right]'\widehat{W}_{A,T}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial f(x_t,\alpha)}{\partial\alpha'}\right] \qquad \text{(A63)}\\
&\qquad - 2\left\{\left[\frac{1}{T}\sum_{t=1}^{T} f(x_t,\alpha)\right]'\widehat{W}_{A,T}\otimes I_{p_{A_s}}\right]\frac{\partial\operatorname{vec}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial f(x_t,\alpha)}{\partial\alpha'}\right]}{\partial\alpha_s'},
\end{aligned}$$

$$\nabla^2_{\alpha_s} q_A(\alpha) = 2F_s(\alpha_s)'W_A F_s(\alpha_s) - 2\left[f_s(\alpha_s)'W_A\otimes I_{p_{A_s}}\right]\frac{\partial\operatorname{vec}\big(F_s(\alpha_s)'\big)}{\partial\alpha_s'}, \tag{A64}$$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial f(x_t,\alpha)}{\partial\alpha'} = F_s(\alpha_s) + o_p(1), \tag{A65}$$

$$\frac{\partial\operatorname{vec}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial f(x_t,\alpha)}{\partial\alpha'}\right]}{\partial\alpha_s'} = \frac{\partial\operatorname{vec}\big(F_s(\alpha_s)'\big)}{\partial\alpha_s'} + o_p(1), \tag{A66}$$

it follows that

$$\nabla_{\alpha_s}\widehat{q}_A(\alpha) - \nabla_{\alpha_s} q_A(\alpha) = o_p(1), \tag{A67}$$

$$\nabla^2_{\alpha_s}\widehat{q}_A(\alpha) - \nabla^2_{\alpha_s} q_A(\alpha) = o_p(1), \tag{A68}$$

uniformly in $\alpha_s \in \mathbf{A}_s$ as in Assumption 1(b). Because

$$\nabla^2_{\alpha_s} q_A(\alpha_0) = 2F_s(\alpha_{s,0})'W_A F_s(\alpha_{s,0}) - 2\left[f_s(\alpha_{s,0})'W_A\otimes I_{p_{A_s}}\right]\frac{\partial\operatorname{vec}\big(F_s(\alpha_{s,0})'\big)}{\partial\alpha_{s,0}'}, \tag{A69}$$

Assumption 5(h) implies Assumption 2(c). Because

$$\begin{aligned}
\widehat{q}_{A,T}(\alpha) &= \left[\frac{1}{T}\sum_{t=1}^{T} f(x_t,\alpha)\right]'\widehat{W}_{A,T}\left[\frac{1}{T}\sum_{t=1}^{T} f(x_t,\alpha)\right]\\
&= \left[f_s(\alpha_s) + T^{-\frac{1}{2}} f_w(\alpha) + O_p\big(T^{-\frac{1}{2}}\big)\right]'W_A \qquad \text{(A70)}\\
&\quad\times\left[f_s(\alpha_s) + T^{-\frac{1}{2}} f_w(\alpha) + O_p\big(T^{-\frac{1}{2}}\big)\right]\\
&= f_s(\alpha_s)'W_A f_s(\alpha_s) + O_p\big(T^{-1}\big)
\end{aligned}$$

uniformly in $\alpha \in \mathbf{A}$, $q_A(\alpha)$ can be written as $q_{A_s}(\alpha_s) + T^{-1/2} q_{A_w}(\alpha)$ where $q_{A_s}(\alpha_s) = (\gamma - f_s(\alpha_s))' W_A(\gamma - f_s(\alpha_s))$ and $q_{A_w}(\alpha) = O_p(1)$ uniformly in $\alpha \in \mathbf{A}$ as in Assumption 2(a). Because W_A is positive definite, the uniqueness of α_s imply Assumption 2(b). □

Proof of Proposition 2(b). Because

$$\nabla_{\alpha_s} \widehat{q}_{A,T}(\alpha) = 2\left[\frac{1}{T}\sum_{t=1}^{T} \frac{\partial f(x_t, \alpha)}{\partial \alpha'}\right]' \widehat{W}_{A,T} \frac{1}{T}\sum_{t=1}^{T} f(x_t, \alpha), \tag{A71}$$

$$\nabla_{\alpha_s} q_A(\alpha) = 2F_s(\alpha_s)' W_A\big(f_s(\alpha_s)\big), \tag{A72}$$

it follows from the compactness of $\mathbf{A}$, the three times continuous differentiability of f and $\widehat{W}_{A,T} \xrightarrow{p} W_A$ that

$$\nabla_{\alpha_s} \widehat{q}_{A,T}(\alpha) - \nabla_{\alpha_s} q_A(\alpha) = o_p(1), \tag{A73}$$

uniformly in $\alpha \in \mathbf{A}$ as required in Assumption 1(b). Similarly, it follows that

$$\nabla^2_{\alpha_s} \widehat{q}_A(\alpha) - \nabla^2_{\alpha_s} q_A(\alpha) = o_p(1), \tag{A74}$$

uniformly in $\alpha_s \in \mathbf{A}_s$ as in Assumption 1(b). Assumption 5(h) implies that $\nabla^2_{\alpha_s} q_A(\alpha_0)$ is positive definite as required by Assumption 3(b). Because W is positive definite, the uniqueness of α_s implies Assumption 3(b). □

References

Andrews, D. W. K. (1992), "Generic uniform convergence." *Econometric Theory*, 8, 241–257. [13, 15]

Canova, F. and L. Sala (2009), "Back to square one: Identification issues in DSGE models." *Journal of Monetary Economics*, 56, 431–449. [1, 2]

Chib, S. and I. Jeliazkov (2001), "Marginal likelihood from the Metropolis–Hastings output." *Journal of the American Statistical Association*, 96, 270–281. [3, 4, 8]

Geweke, J. (1999), "Using simulation methods for Bayesian econometric models: inference, development, and communication." *Econometric Reviews*, 18, 1–73. [2, 4, 8]

Guerron-Quintana, P., A. Inoue, and L. Kilian (2017), "Impulse response matching estimators for DSGE models." *Journal of Econometrics*, 196, 144–155. [5, 6, 7]

Hong, H. and B. Preston (2012), "Bayesian averaging, prediction and nonnested model selection." *Journal of Econometrics*, 167, 358–369. [3, 4]

Inoue, A. and Shintani M. (2018), "Quasi-Bayesian model selection." *Quantitative Economics*, 9, 1265–1297. [1, 7]

Sims, C. A., D. F. Waggoner, and T. Zha (2008), "Methods for inference in large multiple-equation Markov-switching models." *Journal of Econometrics*, 146, 255–274. [2, 4, 8]

Co-editor Frank Schorfheide handled this manuscript.

Manuscript received 23 June, 2015; final version accepted 16 February, 2018; available online 12 March, 2018.