Econometric Theory, **37**, 2021, 708–746.
doi:10.1017/S0266466620000286

LARGE SAMPLE PROPERTIES OF BAYESIAN ESTIMATION OF SPATIAL ECONOMETRIC MODELS*

XIAOYI HAN
The Wang Yanan Institute for Studies in Economics, Xiamen University
LUNG-FEI LEE
The Ohio State University
XINGBAI XU
Xiamen University

This paper studies asymptotic properties of a posterior probability density and Bayesian estimators of spatial econometric models in the classical statistical framework. We focus on the high-order spatial autoregressive model with spatial autoregressive disturbance terms, due to a computational advantage of Bayesian estimation. We also study the asymptotic properties of Bayesian estimation of the spatial autoregressive Tobit model, as an example of nonlinear spatial models. Simulation studies show that even when the sample size is small or moderate, the posterior distribution of parameters is well approximated by a normal distribution, and Bayesian estimators have satisfactory performance, as classical large sample theory predicts.

1. INTRODUCTION

Bayesian estimation has been widely studied in spatial econometrics. Researchers have focused on computation, simulation, and empirical applications of the Bayesian approach to spatial econometric models. Several spatial econometric packages or toolboxes, for example, the spatial econometric toolbox in MATLAB and the package spatialprobit in R, utilize Bayesian estimation. Among them, LeSage and Pace (2009) study Markov Chain Monte Carlo (MCMC) for Bayesian estimation of a spatial autoregressive (SAR) model, the matrix exponential spatial specification model, the latent SAR spatial Tobit model, and the latent-ordered

*We are grateful to the editor, the co-editor, and four anonymous referees for their helpful comments. X.H. gratefully acknowledges the financial support of the Chinese Natural Science Fund (Nos. 71501163 and 71973113), and the Fundamental Research Funds for the Central Universities (20720151144) to Xiamen University. X.X. gratefully acknowledges the financial support of the Fundamental Research Funds for the Central Universities (Nos. 20720171072 and 20720181052) to Xiamen University, and the Chinese Natural Science Fund (Nos. 71703135 and 11771361). We also thank the help from IT Engineer Xiang Li at the School of Economics, Xiamen University. Address correspondence to Xingbai Xu, MOE Key Laboratory of Econometrics, Fujian Key Laboratory of Statistical Science, The Wang Yanan Institute for Studies in Economics, Department of Statistics, School of Economics, Xiamen University, Xiamen, China; e-mail: xu.510@buckeyemail.osu.edu.

© Cambridge University Press 2020

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

spatial probit model. Han and Lee (2016) investigate Bayesian estimation of spatial panel autoregressive models with endogenous spatial weights matrices. Han, Hsieh, and Lee (2017) examine Bayesian estimation and model selection of high-order SAR models.

As Bayesian estimation is extensively applied in spatial econometrics, it is desirable and interesting to understand asymptotic properties of posterior probability density functions (pdf) and Bayesian estimators (BEs) for spatial models from the classical statistics point of view, which is the primary purpose of this paper. Moreover, we examine through simulation studies the finite sample performance of the classical asymptotic approximation of the posterior distribution and asymptotic properties of the BE and, more generally, the quasi-Bayesian estimator (QBE).

There is a literature that studies classical asymptotic properties of posterior pdf and BEs. The Bernstein–von Mises Theorem states that, for independently and identically distributed (i.i.d.) data, the posterior pdf behaves like a normal distribution (Thm. 8.2 in Lehmann and Casella, 1998). Various studies have extended this result to different settings. Chen (1985) establishes three basic conditions for the asymptotic normality of the limiting pdf. Phillips and Ploberger (1996) extend the Bernstein–von Mises Theorem to the time series case. Phillips (1996) investigates the large sample theory of Bayesian model selection in time series models. Kim (1998) establishes classical large sample properties of posterior pdfs of nonstationary time series models. Chernozhukov and Hong (CH hereafter) (2003) generalize the Bernstein–von Mises Theorem to quasi-Bayesian estimation, which is defined similarly as Bayesian estimation and includes general extremum estimation approaches such as generalized method of moments(GMM). Creel et al. (2016) apply the approximate Bayesian computation (ABC) to study GMM estimators and establish classical asymptotic properties of the ABC GMM estimator. Cheng, Gao, and Phillips (2018) investigate classical asymptotic properties of a posterior conditional mean estimator.

CH (2003) provides a general framework for the asymptotic distribution of a posterior probability density and a BE. Our paper relies on the general theory in CH (2003) to study the large sample theory for spatial econometric models.

In Section 2, we introduce a BE of the high-order spatial autoregressive model with spatial autoregressive disturbances (SARAR) model and discuss its computational advantages. In Section 3, we establish the Bernstein–von Mises Theorem and the asymptotic distribution of a QBE for the high-order SARAR model and the spatial Tobit model. In Section 4, we introduce the MCMC algorithm for the high-order SARAR model, and verify the validity of normal approximations of the posterior pdf and QBE through Monte Carlo experiments. Section 5 concludes. All proofs, graphs, and tables can be found either in the Appendices or in the Online Supplementary Material.

2. BAYESIAN ESTIMATION OF THE HIGH-ORDER SARAR MODEL

To motivate the discussion on large sample properties of the posterior pdf and BEs of spatial models, we illustrate advantages of the Bayesian method over classical

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

likelihood based inference, using the following high-order SARAR(p,q) model:

$$Y_n = \sum_{j=1}^{p} \lambda_{j0} W_{jn} Y_n + X_n \beta_0 + U_n, \qquad U_n = \sum_{k=1}^{q} \rho_{k0} M_{kn} U_n + V_n, \tag{2.1}$$

where $Y_n = (y_{1,n}, \ldots, y_{n,n})'$ is an $n \times 1$ column vector of dependent variables, $W_{jn} \equiv (w_{rs,jn})$ and $M_{kn} \equiv (m_{rs,kn})$ are nonstochastic $n \times n$ spatial weights matrices, $X_n = (x_{1,n}, \ldots, x_{n,n})'$ is an $n \times K$ matrix of exogenous variables, $U_n = (u_{1,n}, \ldots, u_{n,n})'$ and $V_n = (v_1, \ldots, v_n)'$ are $n \times 1$ column vectors of disturbances. Assume that v_i's are independent $(0, \sigma_0^2)$ and independent of X_n. Denote $\lambda = (\lambda_1, \ldots, \lambda_p)'$, $\rho = (\rho_1, \ldots, \rho_q)'$, $\theta = (\lambda', \beta', \rho', \sigma^2)'$, and the true parameter $\theta_0 = (\lambda_0', \beta_0', \rho_0', \sigma_0^2)'$. Suppose K, p, and q are fixed. Also let $W_n(\lambda) \equiv \sum_{j=1}^{p} \lambda_j W_{jn}$, $W_n \equiv W_n(\lambda_0)$, $M_n(\rho) \equiv \sum_{k=1}^{q} \rho_k M_{kn}$, $M_n \equiv M_n(\rho_0)$, $S_n(\lambda) \equiv I_n - W_n(\lambda)$, $S_n \equiv S_n(\lambda_0)$, $R_n(\rho) \equiv I_n - M_n(\rho)$, and $R_n \equiv R_n(\rho_0)$. The quasi-likelihood (QL) function for this model is

$$L_n(\theta|Y_n) = (2\pi\sigma^2)^{-n/2} |R_n(\rho)| \cdot |S_n(\lambda)| \exp\left\{ -\frac{\|R_n(\rho)[S_n(\lambda)Y_n - X_n\beta]\|_2^2}{2\sigma^2} \right\}, \tag{2.2}$$

where $||\cdot||_2$ is the Euclidean norm of a vector.

Estimating the high-order SARAR(p,q) model by the quasi-maximum likelihood (QML) is more challenging than a standard SAR model with a single spatial weights matrix.[1] There are two reasons. On one hand, one needs to perform a constrained optimization for equation (2.2) on multidimensional spatial parameters λ and ρ, which satisfy the stability condition $||W_n(\lambda)|| < 1$ and $||M_n(\rho)|| < 1$. On the other hand, with a large sample size, the likelihood-based estimation might be computationally demanding because the Jacobian determinants $|S_n(\lambda)|$ and $|R_n(\rho)|$ in the QL function might not be easily evaluated (Lee and Liu, 2010). From Ord (1975), when there is only one spatial weights matrix, $\ln|S_n(\lambda)| = \sum_{i=1}^{n} \ln(1 - \lambda\omega_i)$, where ω_i's are the eigenvalues of the spatial weights matrix W_n. When $p > 1$ or $q > 1$, this trick does not work as one cannot simultaneously decompose several spatial weights matrices by eigenvalue-eigenvector decomposition. One has to calculate the determinants at different values of λ and ρ, which can be time consuming when the sample size is large.

However, LeSage and Pace (2009) show that it is feasible to estimate model (2.1) using MCMC because the stability condition can be directly imposed on the acceptance–rejection step of the Metropolis–Hasting (M–H) algorithm. Moreover, researchers have developed some efficient algorithms to tackle the computational difficulty of $|S_n(\lambda)|$ and $|R_n(\rho)|$ in equation (2.2) in a Bayesian framework. For instance, Han et al. (2017) propose an efficient MCMC sampler for the posterior distribution of parameters in model (2.1), based upon the exchange algorithm

[1] We thank one referee for helping in clarifying the computational difficulty of the likelihood estimation for the high-order SARAR model.

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

of Murray, Ghahramani, and MacKay (2006). They introduce some auxiliary samples into the sampling step of λ and ρ to cancel out $|S_n(\lambda)|$ and $|R_n(\rho)|$ in the expression of the acceptance probability of the M–H acceptance–rejection step. So the direct evaluation of the Jacobian determinants can be avoided. This computational advantage of the Bayesian approach is valuable, but it does leave issues on asymptotic properties of such BEs from the classical statistics point of view. This motivates us to study large sample properties of the posterior distribution and corresponding Bayesian estimates of model (2.1).

Another advantage of the Bayesian method lies in the incorporation of prior information. In the literature on "mixed" estimation for regression models, researchers have found that when one incorporates prior knowledge (restriction) into the sample information, the resulted "mixed" estimates or Bayesian estimates can be better than the least-square estimates in terms of mean squared error (MSE), when the prior restrictions are correct (Theil and Goldberger, 1961; Raiffa and Schlaifer, 1961; Theil, 1963).[2] Similar findings have also been justified in the recent literature on Bayesian penalized regression with shrinkage priors.[3] For the high-order SARAR model (2.1), when X_n contains some highly correlated explanatory variables as well as high orders of their spatial lags, the multicollinearity issue in X_n could be very severe (Pace, LeSage, and Zhu, 2012), which might render the classical QML estimate (QMLE) of β imprecise. Under this scenario, researchers can increase the precision of parameter estimates, in the sense of MSE, by imposing proper restrictions on β through some specific priors. The combination of the prior and the sample information (likelihood) can produce a Bayesian estimate more precise than a QMLE (Shiller, 1973; Han and Lee, 2013) under a finite sample. Therefore, the prior information on β can actually help tackle possible multicollinearity issues among exogenous regressors in X_n in practice. This also motivates us to investigate the asymptotic properties of the posterior distribution and the Bayesian estimates of model (2.1) as a first step, even though finite sample properties remain a topic for further investigation.

[2]If one views priors of parameters as "random restrictions" on parameters themselves, then a Bayesian approach with priors is relevant to a classical approach with restricted parameters. If restrictions are correct and are incorporated into the sample when doing estimation, then root mean squared error (RMSE) would be improved (Theil and Goldberger, 1961; Theil, 1963). Theil and Goldberger (1961) label this procedure as the "mixed" regression estimation. Raiffa and Schlaifer (1961) also investigate the usage of prior information in regression analysis under the Bayesian framework. Their BE turns out to be identical to the mixed estimator in Theil and Goldberger (1961), under the setting of a quadratic loss function (Theil, 1963).

[3]There is a growing literature on Bayesian shrinkage priors for regression models. This literature aims to implement various classical shrinkage methods for regression models, such as lasso, elastic net, and fused lasso, under the Bayesian framework (Park and Casella, 2008; Kyung et al., 2010; Li and Lin, 2010; Van Erp, Oberski, and Mulder, 2019). Instead of adding various penalty terms to the minimization of the sum of squared residuals, this Bayesian literature chooses to impose shrinkage by specifying prior restrictions on parameters. For a linear regression model with a large set of regressors but only a few relevant regressors, the Bayesian shrinkage approach performs similarly to its classical counterpart (Kyung et al., 2010; Van Erp et al., 2019), which has been shown to produce a smaller MSE than the ordinary least squares (OLS) estimator (Hoerl and Kennard, 1970; Tibshirani, 1996).

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

3. CLASSICAL LARGE SAMPLE THEORIES FOR BAYESIAN ESTIMATION OF SPATIAL ECONOMETRIC MODELS

3.1. The Posterior Probability Density Function

For a spatial econometric model, let $\{y_{i,n}, x_{i,n}\}_{i=1}^n$ be the set of sample observations. We assume that there is a true parametric model with a true parameter $\theta_0 \in \Theta \subseteq \mathbb{R}^K$, where Θ is the parameter space and $\mathbb{R}^K$ is the K-dimensional Euclidean space. The log-likelihood function is denoted by $\ln L_n(\theta)$. More generally, $\ln L_n(\theta)$ can be a log QL function, which might not be the correct likelihood function but its corresponding extremum estimator may still be consistent. The QMLE $\hat{\theta}_n$ maximizes $\ln L_n(\theta)$, that is, $\hat{\theta}_n = \arg\max \ln L_n(\theta)$. CH (2003) establish some high level conditions for large sample properties of posterior quasi-pdfs and QBEs. We will apply their results to establish the asymptotic behavior of the likelihood functions and related BEs of a high-order SARAR model and a SAR Tobit model in Sections 3.2–3.4.

Given a prior $\pi(\theta)$ supported on Θ, the posterior pdf or quasi pdf is

$$p_n(\theta) = \frac{L_n(\theta)\pi(\theta)}{\int_\Theta L_n(\phi)\pi(\phi)d\phi}. \tag{3.1}$$

Denote $\Delta_n(\theta) = \frac{\partial \ln L_n(\theta)}{\partial \theta}$, $\Delta_n = \Delta_n(\theta_0)$, $H_n(\theta) \equiv -\frac{1}{n}E\left[\frac{\partial^2 \ln L_n(\theta)}{\partial\theta\partial\theta'}\right]$, $H_n \equiv H_n(\theta_0)$, and $h \equiv \sqrt{n}(\theta - \theta_0) - H_n^{-1}\frac{1}{\sqrt{n}}\Delta_n$. The posterior quasi pdf of h is

$$p_n^*(h) = \frac{1}{\sqrt{n}} p_n\left(\frac{h}{\sqrt{n}} + \theta_0 + H_n^{-1}\frac{1}{n}\Delta_n\right). \tag{3.2}$$

We conclude that for a high-order SARAR model and a SAR Tobit model, the difference between $p_n^*(\cdot)$ and the density of $N(0, H_n^{-1})$ will disappear as $n \to \infty$ under the total variation of κ-moment norm in the next two subsections. For any $\kappa \geqslant 0$, the total variation of κ-moment norm of a function f on a space $S \subseteq \mathbb{R}^K$ is defined as $||f||_{TVM(\kappa)} \equiv \int_S (1 + ||t||_2^\kappa)|f(t)|dt$. Convergence in the total variation of κ-moment norm of the difference between two densities implies convergence in distribution (Van der Vaart, 1998).

For our asymptotic analysis in the classical framework, laws of large numbers and central limit theorems would play important roles. To use such techniques, the log likelihood would usually be expressed in terms of summations over sampling units. For independent sample observations, the log likelihood or the conditional log likelihood function $\ln L_n(\theta)$ can be obviously written in a single summation of individual conditional or unconditional log likelihoods, that is, $\ln L_n(\theta) = \sum_{i=1}^n l_i(\theta)$, where l_i is the log likelihood for the ith sample observation $(y_{i,n}, x_{i,n})$ or the log conditional likelihood of $y_{i,n}$ conditional on $x_{i,n}$. For time series, it is natural to explore a chain decomposition, that is, to decompose the whole log likelihood into the sum of log conditional likelihoods.

For the likelihood of a spatial econometric model, as spatial units do not have a natural ordering, there is no natural chain decomposition. But usually, one

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

may still decompose a log likelihood function of a spatial econometric model of sample size n into a summation of n items. For a linear spatial econometric model such as the SARAR model, as will be shown subsequently, basic statistics in the corresponding log likelihood or QL constructed from normal disturbances are of linear-quadratic forms, which can be written at the true parameter vector in terms of a single summation of martingale difference double arrays. For nonlinear spatial econometric models, even though there might not be an obvious way of decomposition, it might be desirable to write its score vector in terms of a single summation of spatial mixing or spatial near-epoch dependence (NED) processes (Jenish and Prucha, 2009, 2012). In Section 3.3, we will provide a SAR Tobit model to illustrate such a useful decomposition.

3.2. Large Sample Theory for the High-Order SARAR Model

In this section, we study large sample properties of the posterior pdf and some QBEs of the high-order SARAR model in equation (2.1), which includes the SAR model and the spatial error model as two special cases. The SARAR model is widely used in empirical applications. LeSage and Pace (2008) investigate interstate migration flows in the United States using a SAR model with three spatial weights matrices. Lee and Liu (2010) generalize the efficient GMM estimation of the SAR model in Lee (2007) to this model. Gupta and Robinson (2015, 2018) consider the least squares, IV, and QMLE for a SAR model with a slowly increasing number of spatial weights matrices as its sample size increases.

For any $n \times n$ matrix $A = (a_{ij})$, define $||A||_\infty \equiv \sup_{i=1,\ldots,n} \sum_{j=1}^{n} |a_{ij}|$. Let the parameter space Θ be

$$\{(\lambda', \beta', \rho', \sigma^2)' : \sup_n ||W_n(\lambda)||_\infty < 1, \sup_n ||M_n(\rho)||_\infty < 1, \sigma^2 > 0\}. \tag{3.3}$$

Thus, Θ can be written as $\Theta = \Theta_\lambda \times \mathbb{R}^K \times \Theta_\rho \times (0, \infty)$, where $\Theta_\lambda \subset \mathbb{R}^p$ is the parameter space of λ determined by $\sup_n ||W_n(\lambda)||_\infty < 1$ and $\Theta_\rho \subset \mathbb{R}^q$ is the parameter spaces of ρ determined by $\sup_n ||M_n(\rho)||_\infty < 1$ for stability. We assume that the true parameter $\theta_0 \in \Theta$. Here are some observations about Θ: (1) Θ is convex and open. (2) $\lambda = 0 \in \Theta_\lambda$ because $W_n(0) = 0$. Similarly, $\rho = 0 \in \Theta_\rho$. (3) The spectral radii of $W_n(\lambda)$ and $M_n(\rho)$ are less than 1. (4) Θ_λ and Θ_ρ are bounded. As $\sup_n ||W_n(\lambda)||_\infty < 1$ and $\sup_n ||M_n(\rho)||_\infty < 1$, the model is stable. This is a generalization of some standard assumptions in spatial econometric models with a single spatial weights matrix (see, e.g., Lee, 2004 and Jin and Lee, 2013).

We can simplify our subsequent analysis using the concentrated log-likelihood function $\ln L_{cn}(\lambda, \rho)$. Given λ and ρ, when

$$\hat{\beta}_n(\lambda, \rho) = [(R_n(\rho)X_n)'(R_n(\rho)X_n)]^{-1}(R_n(\rho)X_n)'R_n(\rho)S_n(\lambda)Y_n, \tag{3.4}$$

$$\hat{\sigma}_n^2(\lambda, \rho) = \frac{1}{n} \left\| M_{R_n(\rho)X_n} R_n(\rho) S_n(\lambda) Y_n \right\|_2^2, \tag{3.5}$$

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

where $M_{R_n(\rho)X_n} = I_n - (R_n(\rho)X_n)\left[(R_n(\rho)X_n)'(R_n(\rho)X_n)\right]^{-1}(R_n(\rho)X_n)$, the log likelihood function is maximized. As a function of λ and ρ, such a maximized log-likelihood gives the concentrated log likelihood function $\ln L_{cn}(\lambda, \rho)$. Now, we are ready to state our assumptions formally. For a random variable Z, denote $||Z||_{L^p} \equiv (E|Z|^p)^{1/p}$.

Assumption 3.1. $\pi(\theta)$ is continuous at θ_0, bounded, and $\int_\Theta ||\theta||_2^\kappa \pi(\theta)d\theta < \infty$ for some constant $\kappa \geqslant 0$.

Assumption 3.2. (i) v_i's are independent with mean 0 and variance σ_0^2, and $\sup_i ||v_i||_{L^{4+\Delta}} < \infty$ for some $\Delta > 0$;

(ii) X_n is nonstochastic and its elements are uniformly bounded, that is, $\sup_{i,k,n} |x_{ik,n}| < \infty$.

Assumption 3.3. The spatial weights matrices W_{jn} and M_{kn} are nonstochastic and satisfy

$$\sup_n \max_{1\leqslant j\leqslant p, 1\leqslant k\leqslant q} (||W_{jn}||_1, ||M_{kn}||_1, ||W_{jn}||_\infty, ||M_{kn}||_\infty, ||S_n^{-1}||_1, ||R_n^{-1}||_1) < \infty.$$

Assumption 3.1 is from CH (2003), and it is satisfied for a large number of prior probability densities. In Assumption 3.2, $\{x_{i,n}\}$ is assumed to be nonstochastic and uniformly bounded. This assumption is widely utilized in spatial econometrics, see, for example, Jin and Lee (2013), Lee (2004, 2007), and Lee and Liu (2010). We can relax Assumption 3.2(ii) to allow for stochastic and unbounded $x_{i,n}$, but then we need to assume that individuals are located in a Euclidean space, since we need the NED theory in Jenish and Prucha (2012) in this case. We do so in Section 3.3 for the SAR Tobit model. See Assumption 3.9.

Assumption 3.3 is about uniform boundedness of both row and column sums norms for all the spatial weights matrices W_{jn} and M_{kn}. Similar conditions as Assumption 3.3 are widely employed in the literature of spatial econometrics. See, for example, Kelejian and Prucha (1998), Lee (2004, 2007), Lee and Liu (2010), and so forth. Assumptions 3.2 and 3.3 provide foundations for most steps in the proofs, and they imply that $y_{i,n}$ is uniformly $L_{4+\Delta}$ bounded as stated in the following lemma. Recently, some spatial econometric papers attempted to relax this assumption. See Qu and Lee (2015) and Gupta (2019) for discussion of endogenous or stochastic spatial weights matrices.

LEMMA 1. *Under Assumptions 3.2 and 3.3,* $\sup_{i,n} ||y_{i,n}||_{L^{4+\Delta}} < \infty$.

The proof is similar to that for Proposition 1(1) in Xu and Lee (2015) and thus omitted. The next assumption deals with $\partial\Theta$, the boundary of Θ.

Assumption 3.4. (i) Θ_λ and Θ_ρ are bounded sets.

(ii) There is a constant $L_\sigma > 0$ such that

$$\lim_{n\to\infty} \Pr\left[\inf_{\lambda\in\Theta_\lambda, \rho\in\Theta_\rho} \hat{\sigma}_n^2(\lambda, \rho) > L_\sigma\right] = 1. \tag{3.6}$$

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

(iii) There is a constant $L_\beta > 0$ such that

$$\lim_{n\to\infty} \Pr\left[\inf_{\rho\in\Theta_\rho} \min \operatorname{eig} \frac{1}{n}(R_n(\rho)X_n)'(R_n(\rho)X_n) > L_\beta\right] = 1. \tag{3.7}$$

Since we do not require that Θ is compact, we need to deal with the problem of the boundary of Θ. Assumption 3.4(ii) means that effects of disturbance terms in the model are not ignorable. Assumption 3.4(iii) is also employed in Jin and Lee (2013). It implies that there is no multicollinearity between columns of $R_n(\rho)X_n$. If some columns of X_n are eigenvectors of $M_n(\rho)$, Assumption 3.4(iii) will not hold. With Assumption 3.4, the QMLEs for β and σ^2 will be within a compact set with a large probability. This result is summarized in Lemma 2.[4] With a highly dense spatial matrix, it has been shown that the QMLE of a SAR model might occur in the boundary of its parameter space, and the QMLE would not be consistent (Smith, 2009). So our Assumption has also effectively ruled out dense networks.

LEMMA 2. *Under Assumptions 3.2–3.4, there is a bounded set* $\widetilde{\Theta} \equiv \Theta_\lambda \times \widetilde{\Theta}_\beta \times \Theta_\rho \times \widetilde{\Theta}_{\sigma^2} \subseteq \Theta$ *satisfying* $\widetilde{\Theta} \supset B(\theta_0, \gamma_0)$*, where* $\widetilde{\Theta}_\beta$ *and* $\widetilde{\Theta}_{\sigma^2}$ *are compact sets and* $B(\theta_0, \gamma_0)$ *is the open ball centered at* θ_0 *with radius* $\gamma_0 \in (0, d(\theta_0, \partial\Theta))$*, such that*

$$\lim_{n\to\infty} \Pr\left[\arg \sup_{\theta\in\Theta - B(\theta_0,\gamma_0)} \ln L_n(\theta) \in \widetilde{\Theta}\right] = 1. \tag{3.8}$$

Assumption 3.5. (i) The following limiting matrix exists and is positive definite:

$$\lim_{n\to\infty} \frac{1}{n}[X_n, G_{1n}X_n\beta_0, \ldots, G_{pn}X_n\beta_0]'[X_n, G_{1n}X_n\beta_0, \ldots, G_{pn}X_n\beta_0], \tag{3.9}$$

where $G_{ln} \equiv W_{ln}S_n^{-1}$ for $l = 1, \ldots, p$.

(ii) For any $\eta \in (0,1)$, $\sup_{n,\, \rho\in\overline{\eta\Theta_\rho}} \left\|R_n(\rho)^{-1}\right\|_1 < \infty$, where $\overline{\eta\Theta_\rho}$ is the closure of $\eta\Theta_\rho \equiv \{\eta\rho : \rho \in \Theta_\rho\}$.

(iii) For any $\gamma \in (0, \gamma_0)$, where γ_0 is defined in Lemma 2, $vareig[R_n(\rho)R_n^{-1}]'$ $[R_n(\rho)R_n^{-1}]$ is bounded away from 0 uniformly in large enough n and $\rho \in \Theta_\rho - B(\rho_0, \gamma)$,[5] where *vareig* is the sample variance of eigenvalues.

Assumption 3.6. For large enough n, the smallest eigenvalues of $\Omega_n \equiv \frac{1}{n}E\left[\frac{\partial \ln L_n(\theta_0)}{\partial\theta}\frac{\partial \ln L_n(\theta_0)}{\partial\theta'}\right]$ and $H_n \equiv -\frac{1}{n}E\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}$ are uniformly bounded away from 0.

[4] Lemma 2 is essential to verify condition (iii) in Lemma B.2 in Appendix 2.

[5] For any two sets A and B, denote $A - B = \{x \in A : x \notin B\}$.

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

Assumptions 3.5 and 3.6 are related to identification. Assumption 3.5(i) is a regularity condition. It implies that $[X_n, G_{1n}X_n\beta_0, \ldots, G_{pn}X_n\beta_0]$ has full column rank. Let us explain Assumption 3.5(ii). When ρ goes to $\partial\Theta_\rho$, $R_n(\rho)$ might converge to a singular matrix. Thus, $\left\|R_n(\rho)^{-1}\right\|_1$ might go to infinity. Thus, it is too strong to require that $\sup_{n,\,\rho\in\Theta_\rho}\left\|R_n(\rho)^{-1}\right\|_1 < \infty$. On the contrary, we only require that $\left\|R_n(\rho)^{-1}\right\|_1$ is uniformly bounded over compact sets $\overline{\eta\Theta_\rho}$. Similar conditions for the column sum norm are common in the spatial econometrics literature. See, for example, Assump. 7 of Lee (2004), or Assump. 5 in Jin and Lee (2013). With Assumption 3.5, we show in the Appendix that $\frac{1}{n}E\ln L_n(\theta)$ is uniquely maximized at θ_0 for n large enough. But this conclusion is not sufficient for Assumption 3.6, which deals with only the true parameter θ_0.[6] As the expressions for Ω_n and H_n are complicated, it is hard to add more primitive conditions to ensure Assumption 3.6. We assume it holds as done in Lee (2004) and Jin and Lee (2013). The positive definiteness of the Hessian matrix H_n is generally not a necessary condition for identification, thus, it might probably be stronger than required. More discussion about Assumption 3.5 can be found in Appendix A.

THEOREM 1. *Under Assumptions 3.1–3.6, for the SARAR model.*

$$\left\|p_n^*(h) - \bar{p}_n(h)\right\|_{TVM(\kappa)} \xrightarrow{p} 0, \tag{3.10}$$

where $\bar{p}_n(h) = (2\pi)^{-(K+p+q+1)/2}|H_n|^{1/2}\exp\left(-\frac{1}{2}h'H_nh\right)$ *is the pdf of* $N(0, H_n^{-1})$, $p_n^*(h)$ *is defined in equation (3.2), and* κ *is defined in Assumption 3.1.*

As the proof is quite lengthy, we briefly discuss the idea of the proof which is based on Thm. 1 in CH (2003), listed as Lemma B.2 in Appendix B. With Assumption 3.6, it is not hard to verify condition (4) in Lemma B.2. The difficult part is to verify condition (3) in Lemma B.2, which relaxes the compactness assumption of Θ. To verify condition (3) in Lemma B.2, by Lemma 2, it suffices to show that for any $\delta > 0$, there exists an $\epsilon > 0$, such that

$$\liminf_{n\to\infty}\Pr\left\{\sup_{\theta\in\widetilde{\Theta}:||\theta-\theta_0||_2>\delta}\frac{1}{n}\left[\ln L_n(\theta) - \ln L_n(\theta_0)\right] \leqslant -\epsilon\right\} = 1. \tag{3.11}$$

This condition can be established using Lemma B.3.

(1) For any $\eta \in (0,1)$, denote $\widetilde{\Theta}_\eta \equiv \overline{\eta\Theta_\lambda} \times \widetilde{\Theta}_\beta \times \overline{\eta\Theta_\rho} \times \widetilde{\Theta}_{\sigma^2}$, which is a compact set. We only need to consider those η large enough such that $\theta_0 \in \widetilde{\Theta}_\eta$. Since $\widetilde{\Theta}_\eta$ is compact, it is not hard to show that $\{\frac{1}{n}E\ln L_n(\theta)\}$ is equicontinuous on $\widetilde{\Theta}_\eta$.

(2) Show that for any $\theta \in \Theta$, $\frac{1}{n}E\ln L_n(\theta_0) - \frac{1}{n}E\ln L_n(\theta) \geq 0$.

[6] Although $\frac{1}{n}E\ln L_n(\theta)$ is maximized at θ_0, it is possible that $H_n \equiv -\frac{1}{n}E\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}$ is degenerate.

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

(3) Based on (2), we use Assumption 3.5 to show that for any $\gamma \in (0, \gamma_0)$, where γ_0 is defined in Lemma 2,

$$\liminf_{n\to\infty} \inf_{\theta\in\widetilde{\Theta}:||\theta-\theta_0||_2>\gamma} \frac{1}{n}[E\ln L_n(\theta_0) - E\ln L_n(\theta)] > 0.$$

Then condition (i) in Lemma B.3 holds.

(4) By Thm. 1 in Andrews (1992), we can show $\sup_{\theta\in\widetilde{\Theta}} \frac{1}{n}\|\ln L_n(\theta) - E\ln L_n(\theta)\| = o_p(1)$. Thus, condition (ii) of Lemma B.3 is satisfied. And by Lemma B.3, condition (3) in Lemma B.2 holds.

3.3. The SAR Tobit Model

To investigate large sample properties of the posterior probability density and BE of a nonlinear spatial econometric model, we investigate the SAR Tobit model in this section, which is formulated as

$$y_{i,n} = \max(0, \lambda_0 w_{i\cdot,n} Y_n + x'_{i,n}\beta_0 + \epsilon_i), \tag{3.12}$$

where $w_{i\cdot,n}$ is the ith row of the nonstochastic $n \times n$ spatial weights matrix $W_n \equiv (w_{ij,n})$, $x_{i,n}$ is a K dimensional column vector of exogenous variables, and $\{\epsilon_i\}$ is the disturbance term. In matrix form, $Y_n = \max(0, \lambda_0 W_n Y_n + X_n\beta_0 + \epsilon)$, where $\max(0, U_n) \equiv (\max(0, u_1), \ldots, \max(0, u_n))'$ for any vector $U_n = (u_1, \ldots, u_n)'$. Assume that ϵ_i's are *i.i.d.* $N(0, \sigma^2)$, and $\epsilon = (\epsilon_1, \ldots, \epsilon_n)'$ and $X_n = (x_{1,n}, \ldots, x_{n,n})'$ are independent. The MLE of this model has been considered in Xu and Lee (2015). Denote $z_{i,n}(\theta) \equiv (y_{i,n} - \lambda w_{i\cdot,n} Y_n - x'_{i,n}\beta)/\sigma$. The log-likelihood function is

$$\begin{aligned}\ln L_n(\theta) = &\sum_{i=1}^{n} 1(y_{i,n} = 0) \ln \Phi(z_{i,n}(\theta)) - \frac{1}{2}\ln(2\pi\sigma^2)\sum_{i=1}^{n} 1(y_{i,n} > 0) \\ &+ \ln|I_{2,n} - \lambda W_{22,n}| - \frac{1}{2}\sum_{i=1}^{n} 1(y_{i,n} > 0) z^2_{i,n}(\theta),\end{aligned} \tag{3.13}$$

where $\theta = (\lambda, \beta', \sigma^2)' \in \mathbb{R}^{k+2}$ is the parameter, $W_{22,n}$ is the principal submatrix of W_n corresponding to the strictly positive $y_{i,n}$'s, $|A|$ is the determinant of a square matrix A, and $I_{2,n}$ is the identity matrix with the same dimension as $W_{22,n}$. Notice that in equation (3.13), $\ln L_n(\theta)$ is not directly in summation form. Denote $G_n(Y_n) = \text{diag}(1(y_{1,n} > 0), \ldots, 1(y_{n,n} > 0))$ and $\widetilde{W}_n = G_n(Y_n) W_n G_n(Y_n)$. By a Taylor series expansion,

$$\ln|I_{2,n} - \lambda W_{22,n}| = -\sum_{l=1}^{\infty} \frac{\lambda^l}{l} tr(\widetilde{W}_n^{\,l}) = -\sum_{i=1}^{n}\left[\sum_{l=1}^{\infty} \frac{\lambda^l}{l}(\widetilde{W}_n^{\,l})_{ii}\right]. \tag{3.14}$$

Thus, $\ln L_n(\theta) = \sum_{i=1}^{n} l_{i,n}(\theta)$, where

$$l_{i,n}(\theta) = 1(y_{i,n} = 0) \ln \Phi(z_{i,n}(\theta)) - \frac{1}{2}[\ln(2\pi\sigma^2) + z^2_{i,n}(\theta)] 1(y_{i,n} > 0)$$

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

$$-\sum_{l=1}^{\infty}\frac{\lambda^l}{l}(\widetilde{W}_n^{\,l})_{ii}. \tag{3.15}$$

We add similar assumptions as in Xu and Lee (2015) such that the tools in CH (2003) can be applied.

Assumption 3.7. The parameter space Θ for the SAR Tobit model is a convex and compact subset of $\mathbb{R}^{K+2}$; $\theta_0 \in \Theta^0$, where Θ^0 is the interior of Θ.

Assumption 3.8. $\pi(\theta)$ is continuous at θ_0, bounded, and $\int_\Theta ||\theta||_2^\kappa \pi(\theta)d\theta < \infty$ for some constant $\kappa \geqslant 0$.

Assumption 3.9. Individuals[7] studied in the economy are located in $\mathbb{R}^d$ for some positive integer d. The Euclidean distance d_{ij} between any two different individuals i and j satisfies $d_{ij} \geqslant 1$.

Assumption 3.10. $\zeta \equiv \lambda_m \sup_n ||W_n||_\infty < 1$, and $\Lambda = [-\lambda_m, \lambda_m]$ is the compact parameter space of λ on the real line.

Assumption 3.11. In addition to $w_{ii,n} = 0$ for all i, the weights $w_{ij,n}$ in W_n satisfy at least one of the following two conditions:

(1) When $d_{ij} > \bar{d}_0$, where $\bar{d}_0 > 1$ is a constant, $w_{ij,n} = 0$.
(2) (i) For every n, the number of columns, $w_{\cdot j,n}$, of W_n with $|\lambda_0|\sum_{i=1}^n |w_{ij,n}| > \zeta$, is less than or equal to some fixed nonnegative integer that does not depend on n, denoted as N; (ii) there exists an $\alpha > d$ and a constant C_0 such that $|w_{ij,n}| \leqslant C_0/d_{ij}^\alpha$.

Assumption 3.12. (i) X_n and ϵ are independent.
(ii) For each n, ϵ_i's are *i.i.d.* $N(0,\sigma^2)$ random variables.
(iii) $\sup_{i,k,n} ||x_{ik,n}||_{8+\delta} < \infty$ for some $\delta > 0$.
(iv) $\{x_{i,n}\}_{i=1}^n$ is an α-mixing random field with α-mixing coefficient $\alpha(u,v,r) \leqslant (u+v)^\tau \hat{\alpha}(r)$ for some $\tau \geqslant 0$, where $\hat{\alpha}(r)$ satisfies $\sum_{r=1}^\infty r^{d(\tau_*+1)-1}\hat{\alpha}(r)^{\tilde{\delta}/(4+2\tilde{\delta})} < \infty$ for some $\tilde{\delta} \in (0, 2+\frac{\delta}{2})$, where $\tau_* = \tilde{\delta}\tau/(2+\tilde{\delta})$.

Assumption 3.13. $\limsup_{n\to\infty} \frac{1}{n}[\mathrm{E}\ln L_n(\theta) - \mathrm{E}\ln L_n(\theta_0)] < 0$ for any $\theta \neq \theta_0$.

Assumption 3.14. $\Sigma_0 = \lim_{n\to\infty}\Sigma_n$ exists and is nonsingular, where $\Sigma_n = \frac{1}{n}\mathrm{Var}\sum_{i=1}^n q_{i,n}(\theta_0)$.

Since the SAR Tobit model is a nonlinear spatial model, so far we have not yet known how to relax Assumption 3.7 to allow for a noncompact parameter space. Without compactness, the analysis is more complex since we would have to deal with uniform convergence of the average log likelihood function over a

[7]Here, individuals can be companies, cities, counties, and so forth. We regard each individual as a point in $\mathbb{R}^d$. In graph theory, an individual is often called a node or a vertex.

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

noncompact parameter space. Assumptions 3.9–3.13 are discussed in detail in Xu and Lee (2015). With these conditions, Xu and Lee (2015) established the consistency and asymptotic normality of the MLE of the SAR Tobit model. In their proofs, the conditions of Lemma B.2 are verified. Thus, the posterior pdf of the SAR Tobit model can also be approximated by a normal density.

THEOREM 2. *Under Assumptions 3.7–3.14, for the SAR Tobit model,*

$$\left\|p_n^*(h) - \bar{p}_n(h)\right\|_{TVM(\kappa)} \xrightarrow{p} 0, \tag{3.16}$$

where $\bar{p}_n(h) = (2\pi)^{-(K+2)/2}|H_n|^{1/2}\exp\left(-\frac{1}{2}h'H_nh\right)$ *is the pdf of* $N(0, H_n^{-1})$, $p_n^*(h)$ *is defined in equation (3.2), and* κ *is defined in Assumption 3.8.*

3.4. The Asymptotic Properties of QBE and Confidence Intervals

Given the result of the asymptotic normality of posterior pdf in the high-order SARAR model and the SAR Tobit model, the analysis for asymptotic properties of a BE or a QBE will be similar to that in CH (2003). Given a loss function $\rho_n(\cdot)$, we have a QBE: $\theta_n^* = \arg\min_{\zeta\in\Theta}\int_\Theta \rho_n(\theta-\zeta)p_n(\theta)d\theta$. Let the dimension of θ be $\bar{K}$, where $\bar{K} = K+p+q+1$ in the high-order SARAR model and $\bar{K} = K+2$ in the SAR Tobit model. We are interested in the following three loss functions. When $\rho_n(u) \equiv nu'u$, θ_n^* is the quasi-posterior mean; when $\rho_n(u) = \sqrt{n}\sum_{j=1}^{\bar{K}}|u_j|$, θ_n^* is the quasi-posterior median; when $\rho_n(u) = \sqrt{n}\sum_{j=1}^{\bar{K}}[\tau_j - 1(u_j \leqslant 0)]u_j$, where $\tau_j \in (0,1)$, $\theta_{j,n}^*$ is the quasi-posterior τ_jth quantile. The following regularity conditions in Assumption 3.15 on $\rho_n(\cdot)$ hold for all the above three loss functions (CH, 2003).

Assumption 3.15. The loss function $\rho_n(\cdot)$ has the representation $\rho_n(u) = \rho_1(\sqrt{n}u)$ for some function ρ_1, which satisfies the conditions (i) $\rho_1(u) = 0$ iff $u = 0$; (ii) $\rho_1(\cdot)$ is convex, and $\rho_1(h) \leqslant 1 + ||h||_2^\kappa$, where $\kappa \geqslant 1$ is defined in Assumption 3.1 or 3.8; and (iii) $\varphi(\xi) \equiv \int_{\mathbb{R}^{\bar{K}}}\rho_1(u-\xi)e^{-u'Au}du$ is minimized uniquely at some $\xi^* \in \mathbb{R}^{\bar{K}}$ for any positive definite matrix A.

Denote $\xi_n \equiv \arg\inf_{z\in\mathbb{R}^{\bar{K}}}\int_{\mathbb{R}^{\bar{K}}}\rho_n(u-z)\bar{p}_n(u)du$, where $\bar{p}_n(u)$ is the pdf of $N(0, H_n^{-1})$ in Theorem 1 or 2. When $\rho_1(h) = \rho_1(-h)$, $\xi_n = 0$ (CH, 2003). When $\rho_n(h) = \sqrt{n}\sum_{j=1}^{\bar{K}}[\tau_j - 1(h_j \leqslant 0)]h_j$, ξ_n is the τth quantile of $\bar{p}_n(\cdot)$. In the proof of Theorems 1 and 2, we have verified Assump. 3 and 4 in CH (2003). Thus, by Thm. 2 in CH (2003), we have the following theorem.

THEOREM 3. *Suppose the assumptions in Theorem 1 or 2 hold. Then under Assumption 3.15,*

$$\Omega_n^{-1/2}H_n\left[\sqrt{n}(\theta_n^* - \theta_0) - \xi_n\right] \xrightarrow{d} N(0, I_{\bar{K}}). \tag{3.17}$$

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

When $\Omega_n = H_n$ in the case of MLE, the distribution of $\sqrt{n}(\theta_n^* - \theta_0) - \xi_n$ is approximately $N(0, H_n^{-1})$, the same as that of $\bar{p}_n(\cdot)$. In general, θ_n^* has the same asymptotic distribution as the QMLE, whose asymptotic distribution has a sandwich form. Thus, the posterior distribution of h provides an approximation for $\sqrt{n}(\theta_n^* - \theta_0) - \xi_n$.

Next, consider confidence intervals of a quantity, say, $g(\theta_0)$. Denote the quasi posterior cdf of $g(\theta)$ as $F_{g,n}(a) = \int_\Theta 1(g(\theta) \leqslant a) p_n(\theta) d\theta$, and the corresponding quasi posterior quantile $c_{g,n}(a) = \inf\{b : F_{g,n}(b) \geqslant a\}$ with $a \in (0,1)$. A natural confidence interval is $[c_{g,n}(\frac{a}{2}), c_{g,n}(1-\frac{a}{2})]$. After drawing an MCMC sequence $(\theta^{(1)}, \ldots, \theta^{(B)})$ from the posterior distribution $p_n(\theta)$,[8] we can estimate $c_{g,n}(\frac{a}{2})$ and $c_{g,n}(1-\frac{a}{2})$ from the empirical distribution constructed by $g(\theta^{(1)}), \ldots, g(\theta^{(B-1)}), g(\theta^{(B)})$. When $\Omega_n = H_n$, this interval is asymptotically equal to the confidence interval constructed by the delta method:

$$\left[g(\hat{\theta}) + q_{a/2} \sqrt{\frac{1}{n} \frac{\partial g(\hat{\theta})}{\partial \theta'} \hat{H}_n^{-1} \frac{\partial g(\hat{\theta})}{\partial \theta}}, g(\hat{\theta}) + q_{1-a/2} \sqrt{\frac{1}{n} \frac{\partial g(\hat{\theta})}{\partial \theta'} \hat{H}_n^{-1} \frac{\partial g(\hat{\theta})}{\partial \theta}} \right], \quad \textbf{(3.18)}$$

where $\hat{H}_n^{-1} = H_n^{-1} + o_p(1)$ and q_a is the a-quantile of $N(0,1)$. Thus, $[c_{g,n}(\frac{a}{2}), c_{g,n}(1-\frac{a}{2})]$ is a valid confidence interval.

When $\Omega_n \neq H_n$, as $H_n^{-1} \neq H_n^{-1} \Omega_n H_n^{-1}$, the above approach should be adjusted. By Theorem 3, the variance matrix of the posterior mean is the same as that of the QMLE, approximately equal to $H_n^{-1} \Omega_n H_n^{-1}$, instead of H_n^{-1}. Thus, we can construct a confidence interval for $g(\theta_0)$ based on the asymptotic variance of the QMLE:

$$\left[g(\hat{\theta}) + q_{a/2} \sqrt{\frac{1}{n} \frac{\partial g(\hat{\theta})}{\partial \theta'} \hat{H}_n^{-1} \hat{\Omega}_n \hat{H}_n^{-1} \frac{\partial g(\hat{\theta})}{\partial \theta}}, g(\hat{\theta}) \right.$$
$$\left. + q_{1-a/2} \sqrt{\frac{1}{n} \frac{\partial g(\hat{\theta})}{\partial \theta'} \hat{H}_n^{-1} \hat{\Omega}_n \hat{H}_n^{-1} \frac{\partial g(\hat{\theta})}{\partial \theta}} \right], \quad \textbf{(3.19)}$$

where $\hat{\Omega}_n$ is a consistent estimator of Ω_n.

4. SIMULATION STUDY

In this section, we conduct some simulation studies to analyze the finite sample behavior of the posterior densities for the SARAR and SAR Tobit models, and the finite sample distributions of corresponding BEs. We report additional simulation results on quasi-posterior densities and empirical densities of quasi-Bayesian estimates for the SARAR model in the supplementary appendices. In the following subsections, we discuss the Bayesian MCMC estimation algorithm for the high-

[8] We will discuss an MCMC algorithm for the high-order SARAR model in Section 4.1

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

order SARAR model (2.1),[9] introduce the Monte Carlo simulation designs, and present the simulation results.

4.1. The MCMC Sampler of the SARAR Model

As in Section 3.2, a sufficient condition to ensure the reduced form of equation (2.1) exists is $||W_n(\lambda)||_\infty < 1$ and $||M_n(\rho)||_\infty < 1$. This condition can be imposed on the sampling step of λ and ρ. However, one drawback of equation (3.3) as the stability condition is that it might be computationally demanding with a large sample. This is because the calculation of the two maximum row sum norms would involve a lot of elementary operations for large n. And these two norms need to be evaluated in the acceptance–rejection M–H step at each iteration. To simplify computation, we consider an alternative stronger stability condition for λ and ρ. As $||\sum_{j=1}^{p}\lambda_j W_{jn}||_\infty \leq (\sum_{j=1}^{p}|\lambda_j|)\max_{j=1,\ldots,p}||W_{jn}||_\infty$ and $||\sum_{k=1}^{q}\rho_k M_{kn}||_\infty \leq (\sum_{k=1}^{q}|\rho_k|)\max_{k=1,\ldots,q}||M_{kn}||_\infty$, when

$$\left(\sum_{j=1}^{p}|\lambda_j|\right)\max_{j=1,\ldots,p}||W_{jn}||_\infty < 1, \qquad \left(\sum_{k=1}^{q}|\rho_k|\right)\max_{k=1,\ldots,q}||M_{kn}||_\infty < 1. \tag{4.1}$$

Equation (3.3) is satisfied. Since the two norms $\max_{j=1,2,\ldots,p}||W_{jn}||_\infty$ and $\max_{k=1,2,\ldots,q}||M_{kn}||_\infty$ can be evaluated before the MCMC sampler, the corresponding computational burden can be greatly reduced. In the following paragraphs, we describe the Bayesian MCMC algorithm for the high-order SARAR model (2.1), based upon this stronger stability condition. We also provide some Bayesian estimation results for the model, under the broader stability condition implied by equation (3.3), in the supplementary appendices.

Let $\psi = (\lambda', \rho')'$ be the vector of spatial parameters and $\theta = (\psi', \beta', \sigma^2)'$ be the complete parameter vector of model (2.1). Denote the posterior distribution of θ as $p(\theta|Y_n)$ and the likelihood function as $L_n(\theta|Y_n)$. By Bayes' theorem, the posterior distribution of θ can be formulated as[10]

$$p(\theta|Y_n) \propto \pi_\psi(\psi) \times \pi_{\beta|\sigma^2}(\beta|\sigma^2) \times \pi_{\sigma^2}(\sigma^2) \times L_n(\theta|Y_n), \tag{4.2}$$

where $\pi(\cdot)$'s are the prior densities of relevant parameters, $\pi_\psi(\psi) \propto \pi_\lambda(\lambda)\pi_\rho(\rho)$, and

$$L_n(\theta|Y_n) \propto (\sigma^2)^{-\frac{n}{2}} \cdot |R_n(\rho)| \cdot |S_n(\lambda)| \exp\left[-\frac{(S_n(\lambda)Y_n - X_n\beta)'R_n(\rho)'R_n(\rho)(S_n(\lambda)Y_n - X_n\beta)}{2\sigma^2}\right].$$

The priors of θ are specified as:

$$\lambda_j \sim \mathcal{U}(-1,\ 1), j = 1, 2, \ldots, p;\ \rho_k \sim \mathcal{U}(-1,\ 1),\ k = 1, 2, \ldots, q;$$

[9] The MCMC sampler of the SAR Tobit model is provided in Appendix E of the supplementary material. We thank one referee for suggesting considering the SAR Tobit model.

[10] Here the exogenous variable X_n is suppressed from the conditional set of the likelihood function.

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

$$\beta|\sigma^2 \sim N_K\left(\beta_O, \frac{\sigma^2}{\phi} I_K\right),\ \sigma^2 \sim \mathcal{IG}\left(\frac{a}{2}, \frac{b}{2}\right), \tag{4.3}$$

where $\mathcal{U}(-1,\ 1)$ is the uniform distribution on $(-1, 1)$, $N_K(\beta_O, \frac{\sigma^2}{\phi} I_K)$ is the K-dimensional normal distribution with mean β_O and variance matrix $\frac{\sigma^2}{\phi} I_K$, and $\mathcal{IG}(\frac{a}{2}, \frac{b}{2})$ is the inverse gamma distribution with shape parameter $\frac{a}{2}$ and scaling parameter $\frac{b}{2}$. With row-normalized W_{jn}'s and M_{kn}'s in the following simulation studies, we follow LeSage and Pace (2009) to assume uniform priors over the interval $(-1,\ 1)$ for λ_j's and ρ_k's. The previous stronger stability condition, namely, $(\sum_{j=1}^{p} |\lambda_j|) \max_{j=1,2,\ldots,p} ||W_{jn}||_\infty < 1$ and $(\sum_{k=1}^{q} |\rho_k|) \max_{k=1,2,\ldots,q} ||M_{kn}||_\infty < 1$, would be imposed through the M–H sampling step of λ_j's and ρ_k's.[11]

The MCMC sampler consists of three steps. We first apply the adaptive Metropolis (AM) algorithm in Haario, Saksman, and Tamminen (2001) and Roberts and Rosenthal (2009) to sample all spatial parameters in $\psi = (\lambda', \rho')'$. Then two Gibbs steps are applied to sample β and σ^2. Unlike the standard M–H algorithm, which uses the random walk proposal with a covariance matrix being an identity matrix, the AM algorithm uses the historical MCMC draws to construct the covariance matrix of the proposal distribution. At the iteration l, the historical MCMC draws of ψ form $(\psi^{(0)}, \psi^{(1)}, \ldots, \psi^{(l-1)})$. The AM proposal suggested by Haario et al. (2001) and Roberts and Rosenthal (2009) is

$$\begin{aligned}
&\Omega(\psi|\psi^{(0)}, \psi^{(1)}, \ldots, \psi^{(l-1)}) \\
&= \begin{cases} N_{p+q}(\psi^{(l-1)}, 0.1^2 I_{p+q}/(p+q)) & l \le l_0 \\ (1-\xi) N_{p+q}(\psi^{(l-1)}, \frac{2.38^2}{p+q} \eta_{cov}(\psi^{(0)}, \ldots, \psi^{(l-1)})) & \\ \quad + \xi N_{p+q}(\psi^{(l-1)}, \frac{0.1^2}{p+q} I_{p+q}) & l > l_0 \end{cases},
\end{aligned} \tag{4.4}$$

where $\eta_{cov}(\psi^{(0)}, \ldots, \psi^{(l-1)}) = \frac{1}{l} \sum_{s=0}^{l-1} \psi^{(s)} \psi^{(s)'} - \overline{\psi}^{(l-1)} \overline{\psi}^{(l-1)'}$ with $\overline{\psi}^{(l-1)} = \frac{1}{l} \sum_{s=0}^{l-1} \psi^{(s)}$ is an empirical variance matrix, l_0 is the length of the initial sampling period and the scaling factor 2.38^2 optimizes the mixing property of the Metropolis search for the Gaussian proposals (Gelman, Roberts, and Gilks, 1996). In particular, the AM proposal is a mixture of two normal distributions with a ratio parameter ξ when the number of iterations exceeds l_0. The second component in the mixture, $\xi N_{p+q}(\psi^{(l-1)}, \frac{0.1^2}{p+q} I_{p+q})$, prevents us from generating a singular covariance matrix due to some possibly pathological values of $\eta_{cov}(\psi^{(0)}, \psi^{(1)}, \ldots, \psi^{(l-1)})$. In this paper, ξ is set to be 0.05 following Roberts and Rosenthal (2009). In our MCMC sampler, we apply the AM step only to the burn-in draws. After burn-in, we fix the value of the variance–covariance matrix and use a normal distribution proposal (with that variance–covariance matrix) to

[11] The sampling step for λ_j's and ρ_k's is an M–H step, so their priors do not play a specific role (not conjugate priors). In particular, those uniform priors would be canceled out in the expression of the acceptance probability. The stability conditions would be imposed through the acceptance–rejection step in the M–H algorithm, not through the priors. For nonrow-normalized W_{jn}'s and M_{kn}'s, one can also impose the previous stability condition in the M–H step.

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

continue the M–H step for ψ. Below is the detailed MCMC algorithm:
Step 1: Sample $\psi = (\lambda', \rho')'$ from $p(\psi|Y_n, \beta, \sigma^2)$.
By Bayes' theorem,

$$p(\psi|Y_n, \beta, \sigma^2) \propto \pi_\psi(\psi) L_n(\theta|Y_n).$$

Let l_b be the length of the burn-in period for the MCMC sampler, where $l_0 < l_b$.
(1.1: Burn-in stage) For $l \leq l_b$, generate a new candidate $\tilde{\psi} = (\tilde{\lambda}', \tilde{\rho}')'$ based upon the AM proposal $\Omega(\psi|\psi^{(0)}, \psi^{(1)}, \ldots, \psi^{(l-1)})$ as in equation (4.4) and check whether $\tilde{\psi}$ satisfies stability conditions in equation (4.1).[12] If not, redraw $\tilde{\psi}$ until it meets those conditions. With the acceptance probability

$$\Pr(\psi^{(l-1)}, \tilde{\psi}) = \min\left\{1, \frac{L(\tilde{\psi}, \beta^{(l-1)}, \sigma^{2(l-1)}|Y_n)}{L(\psi^{(l-1)}, \beta^{(l-1)}, \sigma^{2(l-1)}|Y_n)}\right\},$$

set $\psi^{(l)}$ equal to $\tilde{\psi}$, else set it equal to $\psi^{(l-1)}$.
(1.2: After burn-in) For $l > l_b$, fix the variance–covariance matrix of the proposal to be $Var_\psi = \xi^2 \frac{0.1^2}{p+q} I_{p+q} + (1-\xi)^2 \frac{2.38^2}{p+q} \eta_{cov}(\psi^{(0)}, \psi^{(1)}, \ldots, \psi^{(l_b)})$. Generate $\tilde{\psi}$ from $N_{p+q}(\psi^{(l-1)}, Var_\psi)$ and check whether $\tilde{\psi}$ satisfies equation (4.1). If not, redraw $\tilde{\psi}$ until it meets those conditions. With the acceptance probability specified in Step 1.1, set $\psi^{(l)}$ equal to $\tilde{\psi}$; otherwise, set it equal to $\psi^{(l-1)}$.

Note that we need to evaluate the Jacobian determinants $|R_n(\rho)|$ and $|S_n(\lambda)|$ at each iteration to compute the acceptance probability. This might be computationally intensive when n is large. To simplify computation, one can apply the exchange algorithm in Han et al. (2017) to calculate the acceptance probability. Details of the exchange algorithm to sample λ and ρ are provided in Appendix D of the supplementary appendices.
Step 2: Sample β from $p(\beta|Y_n, \psi, \sigma^2)$.
Denote $Y_n(\lambda, \rho) = R_n(\rho) S_n(\lambda) Y_n$ and $X_n(\rho) = R_n(\rho) X_n$. By Bayes' theorem,

$$\begin{aligned} &p(\beta|Y_n, \psi, \sigma^2) \propto \pi_{\beta|\sigma^2}(\beta|\sigma^2) L_n(\theta|Y_n) \\ &\propto \exp\left(-\frac{\phi}{2\sigma^2}(\beta - \beta_O)'(\beta - \beta_O)\right) \cdot \exp\left[-\frac{(Y_n(\lambda,\rho) - X_n(\rho)\beta)'(Y_n(\lambda,\rho) - X_n(\rho)\beta)}{2\sigma^2}\right] \\ &\sim N_K(T_\beta, \Sigma_\beta), \end{aligned} \tag{4.5}$$

where $\Sigma_\beta = \left[\frac{\phi}{\sigma^2} I_K + \frac{X_n(\rho)' X_n(\rho)}{\sigma^2}\right]^{-1}$ and $T_\beta = \Sigma_\beta \cdot \left[\frac{\phi \beta_O}{\sigma^2} + \frac{X_n(\rho)' Y_n(\lambda,\rho)}{\sigma^2}\right]$. In particular, with $\beta_O = 0$,

$$T_\beta = [\phi I_K + X_n(\rho)' X_n(\rho)]^{-1} X_n(\rho)' Y_n(\lambda, \rho).$$

[12] In the simulation study below, we check whether $(\sum_{j=1}^2 |\tilde{\lambda}_j|) \max_{j=1,2} ||W_{jn}||_\infty < 1$ and $(\sum_{k=1}^2 |\tilde{\rho}_k|) \max_{k=1,2} ||M_{kn}||_\infty < 1$ for the high-order SARAR model. When all spatial weights matrices are row-normalized, the condition becomes the stability condition in Lee and Liu (2010), namely, $\sum_{j=1}^2 |\tilde{\lambda}_j| < 1$ and $\sum_{k=1}^2 |\tilde{\rho}_k| < 1$. We also provide some simulation results in Section G of the supplementary appendices when the M–H step is implemented under the broader stability condition implied by equation (3.3).

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

Conditional on λ and ρ, T_β can be viewed as the ridge regression estimator for β (Hoerl and Kennard, 1970; Marquardt, 1970). The scaling parameter ϕ, which is usually chosen by practitioners to reflect their prior belief on how large the variance of β should be, is the ridge parameter. As ridge regression can achieve large reductions in variances by allowing a small amount of biases (Marquardt, 1970), it has been suggested as one potential solution for multicollinearity among regressors (Congdon, 2006). So we expect the BEs might have better performance than the classical MLEs in terms of MSE.

To implement the Bayesian ridge regression,[13] instead of having ϕ prespecified, one may follow the Bayesian lasso literature (Park and Casella, 2008; Kyung et al., 2010) to assume a hierarchical gamma prior on ϕ, namely, $\phi \sim gamma(c,d)$, and treat it as a hyperparameter to sample.[14] Specifically, we can assume a hierarchical normal prior for β,

$$\beta|\sigma^2 \sim N_K(0, \frac{\sigma^2}{\phi} I_K),\ \phi \sim gamma(c,d).$$

The Gibbs sampling step for ϕ is

$$p(\phi|\beta,\sigma^2) \propto \pi_\phi(\phi)\pi_{\beta|\sigma^2,\phi}(\beta|\sigma^2,\phi)$$
$$\propto \phi^{c-1}d^c \exp(-d\phi) \times \left(\frac{\sigma^2}{\phi}\right)^{-\frac{K}{2}} \times \exp\left(-\frac{\phi\beta'\beta}{2\sigma^2}\right) \sim gamma\left(c+\frac{K}{2}, d+\frac{\beta'\beta}{\sigma^2}\right).$$

Step 3: Sample σ^2 from $p(\sigma^2|Y_n,\psi,\beta)$.
By Bayes' theorem,

$$p(\sigma^2|Y_n,\psi,\beta) \propto \pi_{\sigma^2}(\sigma^2)\pi_{\beta|\sigma^2}(\beta|\sigma^2)L_n(\theta|Y_n)$$
$$\propto (\sigma^2)^{-\frac{a}{2}-\frac{n}{2}} \exp\left(-\frac{b}{2\sigma^2}\right)\left(\frac{\sigma^2}{\phi}\right)^{-\frac{K}{2}} \exp\left[-\frac{\phi(\beta-\beta_O)'(\beta-\beta_O)}{2\sigma^2}\right] \cdot$$
$$\exp\left[-\frac{(Y_n(\lambda,\rho)-X_n(\rho)\beta)'(Y_n(\lambda,\rho)-X_n(\rho)\beta)}{2\sigma^2}\right] \sim \mathcal{IG}\left(\frac{a_p}{2}, \frac{b_p}{2}\right),$$

where $a_p \equiv a+n+K$ and $b_p \equiv b+\phi(\beta-\beta_O)'(\beta-\beta_O)+(Y_n(\lambda,\rho)-X_n(\rho)\beta)'(Y_n(\lambda,\rho)-X_n(\rho)\beta)$.

[13] We provide some simulation results of the Bayesian ridge estimation in Section G of the supplementary appendices.

[14] There are some other ways of determining the value of ϕ in the literature. For instance, Hoerl and Kennard (1970) utilize the ridge trace plot to select ϕ. Golub, Heath, and Wahba (1979) propose a generalized cross-validation approach to determine ϕ. In this paper, we choose to sample ϕ, which is easier to implement with a Gibbs step.

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

4.2. Monte Carlo Simulation Design

Suppose individuals are located in the Euclidean plane $\mathbb{R}^2$ throughout the simulation study. We first consider the high-order SARAR model,

$$Y_n = \lambda_1 W_{1n} Y_n + \lambda_2 W_{2n} Y_n + l_n \beta_1 + \sum_{j=2}^{4} X_{jn} \beta_j + \epsilon_n,\ \epsilon_n$$
$$= \rho_1 M_{1n} \epsilon_n + \rho_2 M_{2n} \epsilon_n + V_n, \tag{4.6}$$

where the intercept term l_n is an $n \times 1$ vector of ones, $X_{2n} = (x_{21}, \ldots, x_{2n})'$, $X_{3n} = (x_{31}, \ldots, x_{3n})'$ and $X_{4n} = (x_{41}, \ldots, x_{4n})'$ are $n \times 1$ column vectors of exogenous regressors, and $V_n \sim N(0, I_n)$. The true parameters are $\lambda_1 = 0.5$, $\lambda_2 = 0.3$, $\rho_1 = 0.4$, $\rho_2 = 0.2$, $\beta_1 = 2$, and $\beta_2 = \beta_3 = \beta_4 = 1$. W_{jn}'s and M_{kn}'s are spatial weights matrices constructed from m nearest neighbors, whose design will be discussed later. We study three sample sizes: $n = 200$, 400, and 800. We explore two different settings for X_{2n}. In the first setting, x_{2i}'s are i.i.d. and follow the Bernoulli(0.5) distribution; in the second setting, $X_{2n} = (I_n - 0.2 \times W_{1n})^{-1} \cdot N_n(0, I_n)$, so x_{2i}'s are spatially correlated. In both settings, X_{3n} and X_{4n} are generated from $N_n(0, 2I_n)$.[15]

For a nonlinear spatial model, we study the SAR Tobit model, namely,[16]

$$y_{i,n} = \max(0,\ \lambda_1 \sum_{j=1}^{n} w_{ij,n} y_{j,n} + x'_{i,n} \beta + v_i), \tag{4.7}$$

where $x_{i,n} = (x_{i1n}, \ldots, x_{i4n})'$, $\beta = (\beta_1, \ldots, \beta_4)'$, $w_{ij,n}$ is the (i,j)th element of the spatial weights matrix W_n, and v_i's are i.i.d. $N(0,1)$. The exogenous regressors $x_{i,n}$'s consist of four regressors and are generated in the same way as those of the SARAR model in equation (4.6). The true values of $\lambda_1 = 0.5$ and β are set to be the same as those for the high-order SARAR model, except that we set the intercept $\beta_1 = -0.5$ to achieve a censoring rate about 0.42. According to Qu and Lee (2013), with a row-normalized W_n, $|\lambda| < 1$ ensures a unique solution to the model given exogenous regressors and disturbances. We follow Qu and Lee (2013) to generate the dependent variable vector $Y_n = (y_{1,n}, \ldots, y_{n,n})'$ based on the iterative contraction mapping algorithm $Y_n^{(j)} = \max(0, \lambda_1 W_n Y_n^{(j-1)} + X_n \beta + V_n)$, starting from $Y_n^{(0)} = X_n \beta + V_n$. The iteration stops when $||Y_n^{(j)} - \max(0, \lambda_1 W_n Y_n^{(j-1)} + X_n \beta + V_n)||_\infty < 10^{-6}$.[17]

All spatial weights matrices in equations (4.6) and (4.7) are constructed from the function "makeneighborsw",[18] which generates a row-normalized spatial weights

[15] We also explore different data generating processes (DGP) for the high-order SARAR model in equation (4.6). The complete simulation designs and results are summarized in Sections F and G of the supplementary appendices.

[16] We thank one referee for suggesting the need of investigating the SAR Tobit model in the simulation study.

[17] There are robust checks on the normality assumption in Xu and Lee (2015). From their simulations, the estimates are not very sensitive to unimodal non-normal distribution, but can be sensitive to bimodal mixture distribution.

[18] This function is taken from James LeSage's MATLAB code for spatial econometrics, which can be found at http://www.spatial-econometrics.com/.

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

matrix based on the m nearest neighbors with various m for different DGP. Taking W_{1n} and M_{1n} as examples, the procedure consists of the following three steps:

1. For each observational unit i on the Euclidean plane $\mathbb{R}^2$, generate its coordinates $xc(i)$ and $yc(i)$ from $\mathcal{X}^2(3)$, the chi-square distribution with three degrees of freedom, to represent its geographical location on $\mathbb{R}^2$.
2. Compute the geographical distance d_{ij} between i and j based on their coordinates; for each i, find the nearest m neighbors based on $d(i,j), j \neq i$ and denote the corresponding $W_{ij1} = 1$ ($M_{ij1} = 1$), otherwise $W_{ij1} = 0$ ($M_{ij1} = 0$) for remaining neighbors.
3. Row-normalize W_{1n} (M_{1n}). For equation (4.6), we set $m = 1$ for W_{1n} and $m = 3$ for M_{1n}. We specify W_{2n} and M_{2n} as, respectively, row-normalized spatial weights matrices based upon only the second and fourth nearest neighbors.[19] These spatial weights matrices are rather sparse. We also explore some denser spatial weights matrices and nonrow-normalized spatial weights matrices in the supplementary appendices. Finally, we set W_n in equation (4.7) for the SAR Tobit model to be a row-normalized spatial weights matrix with $m = 5$. In our Monte Carlo studies, each case has 1,000 repetitions. The generated X_n and W_n are fixed across the 1,000 repetitions in each case, but the disturbances are randomly redrawn in each repetition.[20]

We implement the Bayesian MCMC sampler in Section 4.1, for high-order SARAR models with the exchange algorithm and the SAR Tobit model in Appendices D and E, and utilize the posterior draws of parameters to examine asymptotic properties in Theorems 1–3. We specify priors for parameters of the high-order SARAR model as in equation (4.3). In particular, we set the prior mean of β to be zero to take a neutral stance regarding its sign. The scaling factor $\frac{1}{\phi}$ is set to be 10, which is much larger than the actual variance in the DGPs. The prior of σ^2 is an inverse-gamma prior with the shape parameter $a = 6$ and the scaling parameter $b = 4$. The priors of parameters λ_1, β, and σ^2 in equation (4.7) are specified in a similar manner to equation (4.3).

We make use of Bayesian estimates to justify Theorems 1–3. Regarding the finite sample behavior of the posterior, we first run one Markov chain with 25,000 iterations and obtain the posterior draws of λ, ρ, and β. The first 20% draws

[19] Specifically, we first generate $\tilde{W}_{2n}$($\tilde{M}_{2n}$), which is a row-normalized spatial weights matrix based on 2 (4) nearest neighbors. We then eliminate the first (first, second, and third) nearest neighbors to obtain W_{2n} (M_{2n}).

[20] We compute all pairwise distances associated with W_n for the SAR Tobit model under different sample sizes. The minimum (maximum) pairwise distance is 0.0079 (21.66) for $n = 200$, 0.018 (24.24) for $n = 400$ and 0.0069 (22.53) for $n = 800$, with an average of 4.10 for $n = 200$, 4.16 for $n = 400$ and 4.05 for $n = 800$. Though these cannot be viewed as a formal justification of Assumption 3.9, they do suggest that W_n is more or less compatible with the NED design where all units should be separated by a finite distance.

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

are discarded as burn-in draws.[21] We thereby compare the marginal posterior density plots of parameter draws, demeaned at MLEs, with corresponding normal density plots. For empirical distributions of Bayesian estimates, we first run 1,000 Markov chains (repetitions) with length 25,000 and 20% burn-in ratio, where V_n is regenerated in each repetition. Then we obtain posterior means of λ, ρ, and β in each repetition. Based upon these 1,000 samples, we can plot the empirical densities of posterior means, and compare those plots with the corresponding normal density plots.

We treat the posterior means in each repetition as Bayesian point estimates. We compare not only the density plots, but also the bias and RMSE of the Bayesian estimates with those of MLEs across 1,000 repetitions. In addition, we report the computational time of the MCMC algorithm in Table 2 and the exchange algorithm in Appendix D for the high-order SARAR model in a single Markov chain with 1,000-iteration length, under different values of n.[22] Both algorithms are run on a 3.30 GHz server with an Intel Xeon processor and 8 Gb installed memory.

4.3. Simulation Results

In this subsection, we present simulation results on the marginal posterior densities of some parameters[23] of the high-order SARAR model and the SAR Tobit model. We also summarize the finite sample distributions of the posterior mean estimates for these parameters. More simulation results and discussions under different settings and DGPs are reported in the supplementary appendices of this paper.

We find that even for sample sizes that are not large ($n = 200$), our large sample theories provide a satisfactory approximation for posterior densities and BEs in most cases. As sample sizes increases, such approximations become more precise.

4.3.1. *The high-order SARAR model.* Figure 1 compares plots of marginal posterior densities of λ_j's and β_h's for $j = 1, 2$ and $h = 1, 2$, with plots of their corresponding normal densities, for $n = 200$ under settings 1 and 2. As can be seen from these figures, in most cases, plots of posterior densities and normal densities are already very close to each other. The match between marginal posterior densities and normal densities further improves as n grows to 400 and 800.[24] Figure 2 provides empirical density plots of posterior means of spatial parameters λ_j's and β_k's for $j = 1, 2$ and $k = 1, 2$, with $n = 200$. We see that even when $n = 200$, these empirical densities already behave similarly to their corresponding normal

[21] Some trace plots are depicted in Figures G.1 and G.2 in the supplementary appendices to demonstrate the convergence of the MCMC sampler.

[22] We explore cases of small and moderate sample sizes with $n = 200$, 400, and 800, as well as cases of relatively large sample sizes with $n = 1,000$, 1,200, and $n = 1,400$.

[23] In Figures 1–4, we only report the results for spatial parameters in the main equations, the intercept term β_1, and one or two slope coefficients of the high-order SARAR and the SAR Tobit model, as we are more interested in the main equation and graphs where other parameters are similar. We thank the co-editor for this suggestion.

[24] See Section G.3 of the supplementary appendices for more figures.

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

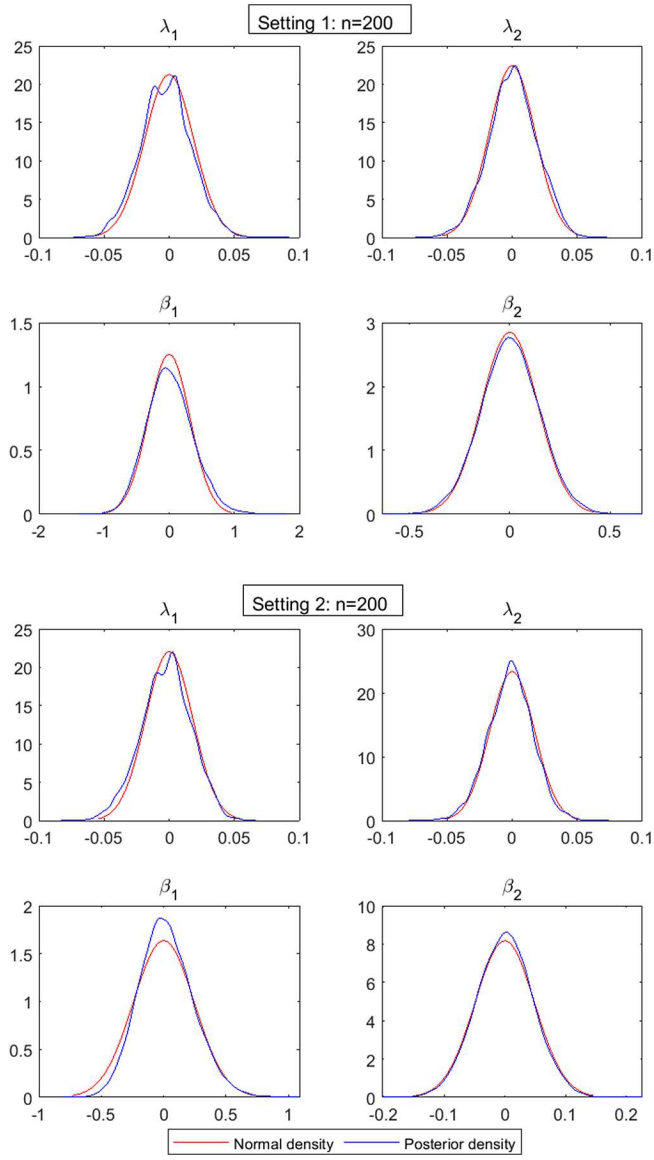


FIGURE 1. Setting 1: x_{2i} i.i.d. Bernoulli(0.5); Setting 2: $X_{2n} = (I_n - 0.2 \times W_{1n})^{-1} \cdot N_n(0, I_n)$. Marginal posterior density versus normal density for the high-order SARAR model: $(\lambda_1, \lambda_2, \beta_1, \beta_2) = (0.5, 0.3, 2, 1)$.

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

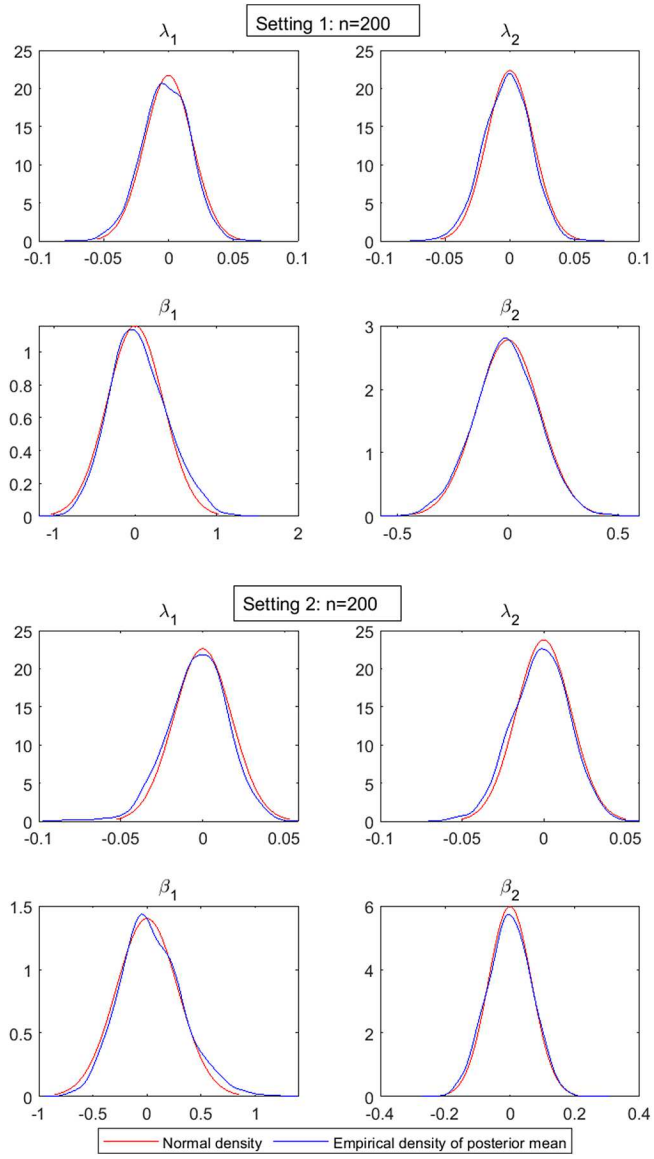


FIGURE 2. Setting 1: x_{2i} i.i.d. Bernoulli(0.5); Setting 2: $X_{2n} = (I_n - 0.2 \times W_{1n})^{-1} \cdot N_n(0, I_n)$. Empirical density of Bayesian posterior mean estimates versus normal density for the high-order SARAR model: $(\lambda_1, \lambda_2, \beta_1, \beta_2) = (0.5, 0.3, 2, 1)$.

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

TABLE 1. Model estimation results of the high-order SARAR model with normal errors

		Setting 1				Setting 2			
		MLE		BE		MLE		BE	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
$n = 200$	λ_1	−0.002	0.018	−0.003	0.018	−0.001	0.018	−0.002	0.018
	λ_2	−0.000	0.018	−0.001	0.018	−0.001	0.017	−0.001	0.017
	ρ_1	−0.012	0.091	−0.005	0.087	−0.011	0.086	−0.004	0.082
	ρ_2	−0.006	0.063	−0.007	0.062	−0.006	0.063	−0.007	0.061
	β_1	0.020	0.363	0.048	0.365	0.015	0.306	0.033	0.303
	β_2	−0.006	0.144	−0.007	0.144	−0.001	0.066	−0.001	0.066
	β_3	−0.001	0.046	−0.002	0.046	−0.002	0.048	−0.003	0.048
	β_4	−0.001	0.048	−0.001	0.048	−0.002	0.049	−0.003	0.049
	σ^2	−0.038	0.109	−0.008	0.103	−0.034	0.106	−0.004	0.101
$n = 400$	λ_1	−0.001	0.013	−0.001	0.013	−0.000	0.012	−0.001	0.012
	λ_2	−0.001	0.014	−0.001	0.014	−0.001	0.012	−0.001	0.012
	ρ_1	−0.005	0.061	−0.001	0.059	−0.007	0.058	−0.003	0.057
	ρ_2	−0.004	0.041	−0.005	0.041	−0.001	0.041	−0.002	0.050
	β_1	0.018	0.249	0.030	0.249	0.011	0.197	0.018	0.196
	β_2	0.004	0.096	0.004	0.096	−0.001	0.050	−0.002	0.050
	β_3	0.001	0.037	0.001	0.037	−0.001	0.032	−0.002	0.032
	β_4	0.001	0.048	0.001	0.034	0.000	0.033	−0.000	0.033
	σ^2	−0.016	0.076	−0.001	0.075	−0.017	0.076	−0.002	0.074
$n = 800$	λ_1	−0.001	0.009	−0.001	0.009	−0.000	0.009	−0.000	0.009
	λ_2	0.000	0.009	−0.000	0.009	−0.000	0.008	−0.000	0.008
	ρ_1	−0.002	0.041	−0.001	0.041	−0.004	0.040	−0.002	0.040
	ρ_2	−0.002	0.029	−0.003	0.029	−0.001	0.029	−0.001	0.029
	β_1	0.011	0.167	0.017	0.166	0.006	0.129	0.010	0.129
	β_2	−0.004	0.069	−0.005	0.069	0.001	0.034	0.001	0.034
	β_3	−0.001	0.024	−0.001	0.024	0.002	0.025	0.001	0.025
	β_4	−0.000	0.024	−0.001	0.024	−0.002	0.026	−0.002	0.026
	σ^2	−0.012	0.053	−0.004	0.052	−0.012	0.054	−0.004	0.053

The values reported in this table are calculated from 1,000 repetitions.
In the 1,000 repetitions of each case, the generated X_n and W_n are fixed across repetitions, but the disturbances are random draws.
Abbreviations: BE, Bayesian estimate; MLE, maximum likelihood estimate; RMSE, root mean squared error.
$(\lambda_1, \lambda_2, \rho_1, \rho_2) = (0.5, 0.3, 0.4, 0.2), (\beta_1, \beta_2, \beta_3, \beta_4, \sigma^2) = (2, 1, 1, 1, 1)$. Setting 1: x_{2i} i.i.d. Bernoulli(0.5); Setting 2:$X_{2n} = (I_n - 0.2 \times W_{1n})^{-1} \cdot N_n(0, I_n)$.

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

TABLE 2. CPU time of Bayesian estimation algorithms: High-order SARAR model with normal errors.

Estimation method		$n = 200$	$n = 400$	$n = 800$	$n = 1,000$	$n = 1,200$	$n = 1,400$
DBE	Setting 1	2.35	26.30	105.75	174.09	256.58	366.43
	Setting 2	2.39	26.34	103.83	170.40	257.85	366.06
EA	Setting 1	2.45	20.60	79.09	127.93	191.41	259.09
	Setting 2	2.34	20.03	79.23	128.20	191.42	260.78

$(\lambda_1, \lambda_2, \rho_1, \rho_2) = (0.5, 0.3, 0.4, 0.2), (\beta_1, \beta_2, \beta_3, \beta_4, \sigma^2) = (2, 1, 1, 1, 1)$.
Abbreviations: DBE, Direct Bayesian Estimation; EA, Exchange Algorithm.
The length of Markov chains for both algorithms: 1,000.
Both algorithms are run on a 3.30 GHz server with an Intel Xeon processor and 8 Gb installed memory.
The CPU times are in seconds.
Setting 1: x_{2i} i.i.d. Bernoulli(0.5); Setting 2: $X_{2n} = (I_n - 0.2 \times W_{1n})^{-1} \cdot N_n(0, I_n)$.

densities. As n grows, these empirical densities become more concentrated and are getting closer to corresponding normal densities.[25]

Table 1 gives biases and RMSEs of MLEs and BEs across 1,000 repetitions under different sample sizes for the high-order SARAR model. We find that biases and RMSEs of the MLEs and Bayesian posterior means are close in most cases. For instance, the biases and RMSEs for MLEs and BEs for λ_1, λ_2, β_2, β_3 are almost the same.[26]

Table 2 reports the time needed to compute the Bayesian MCMC algorithm and the exchange algorithm in a single Markov chain of 1,000 iterations for the high-order SARAR model. When n is relatively small (less than 800), both algorithms can deliver a well-behaved MCMC sample within 20 to 25 s for all cases. In particular, with $n = 200$, the whole Bayesian estimation procedure can be done in 2 s. When n is relatively large (more than 1,000), as the classical ML method might be computationally intensive, the exchange algorithm costs significantly less CPU time than the direct MCMC algorithm. Specifically, it takes the direct MCMC algorithm more than 6 min to finish estimation for the high-order SARAR model

[25] See Section G.4 of the supplementary appendices for more figures.

[26] The co-editor noticed that under the DGP in equation (4.6), the difference between the bias of MLE and BE is small for all parameters except the constant. However, this conclusion does not hold for DGP1 in the supplement. For example, from Table 5 on page 44 of the supplement, for DGP1, under both settings, the differences of the bias for $\hat{\beta}_1$ are small. For DGP1 under setting 2, when $n = 200$, the biases for the MLE and BE are respectively -0.001 and 0.001. When $n = 400$, the biases for both MLE and BE are 0.007.

Notice that in equation (4.6), $\lambda_1 + \lambda_2 = 0.8$ is closer to 1. One possible reason for large biases under equation (4.6) is that, for the MLE and BE of the SAR or SARAR model, when spatial parameter gets closer to the boundary of the parameter space, the bias for the constant term tends to be larger and such change is more sensitive than that of slope parameters. We have done some simulations and our conclusion seems to be correct: fixing λ_1, as λ_2 increases, the bias for β_1 will increase while the biases for other parameters seem to be unchanged; in a SAR model with only one spatial weights matrix, as λ_1 tends to the boundary of the parameter space, the bias for β_1 increases faster than that for slope coefficients.

But this does not affect the large sample theory. We can see that even for equation (4.6), the difference between the biases of MLE and BE for the constant term decreases as n increases. For example, in the setting 2 of Table 1, when $n = 200$, 400, and 800, the differences between the biases are, respectively, 0.018, 0.007, and 0.004.

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

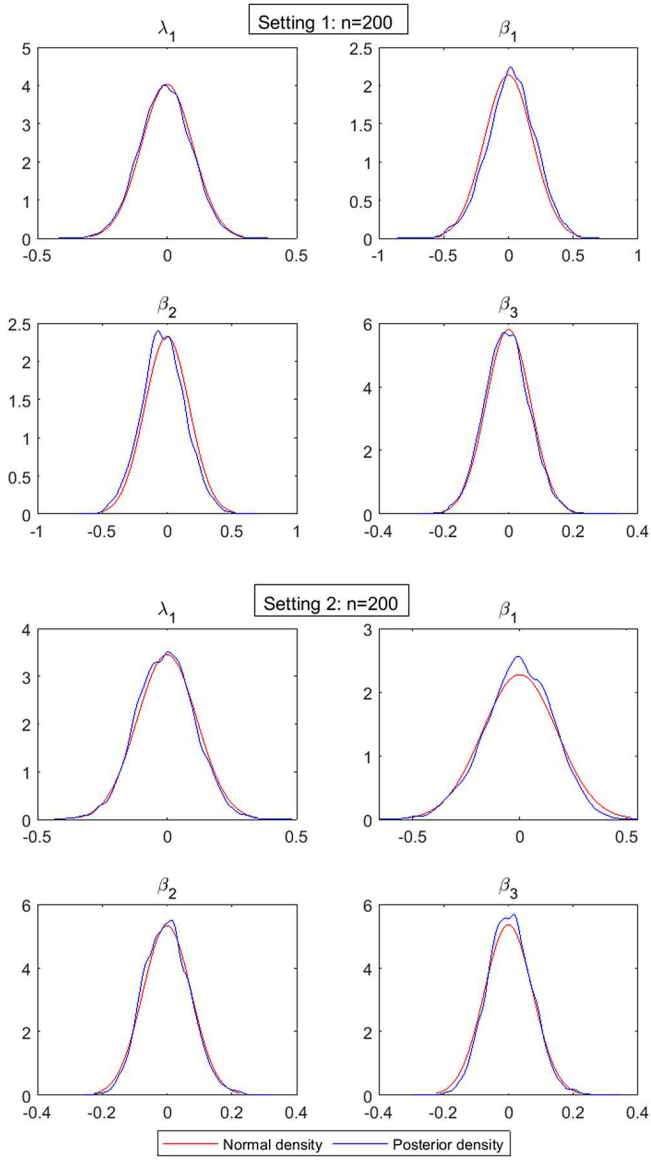


FIGURE 3. Setting 1: x_{2i} i.i.d. Bernoulli(0.5); Setting 2: $X_{2n} = (I_n - 0.2 \times W_{1n})^{-1} \cdot N_n(0, I_n)$. Marginal posterior density versus normal density: the SAR Tobit model with $(\lambda_1, \beta_1, \beta_2, \beta_3) = (0.5, -0.5, 1, 1)$.

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

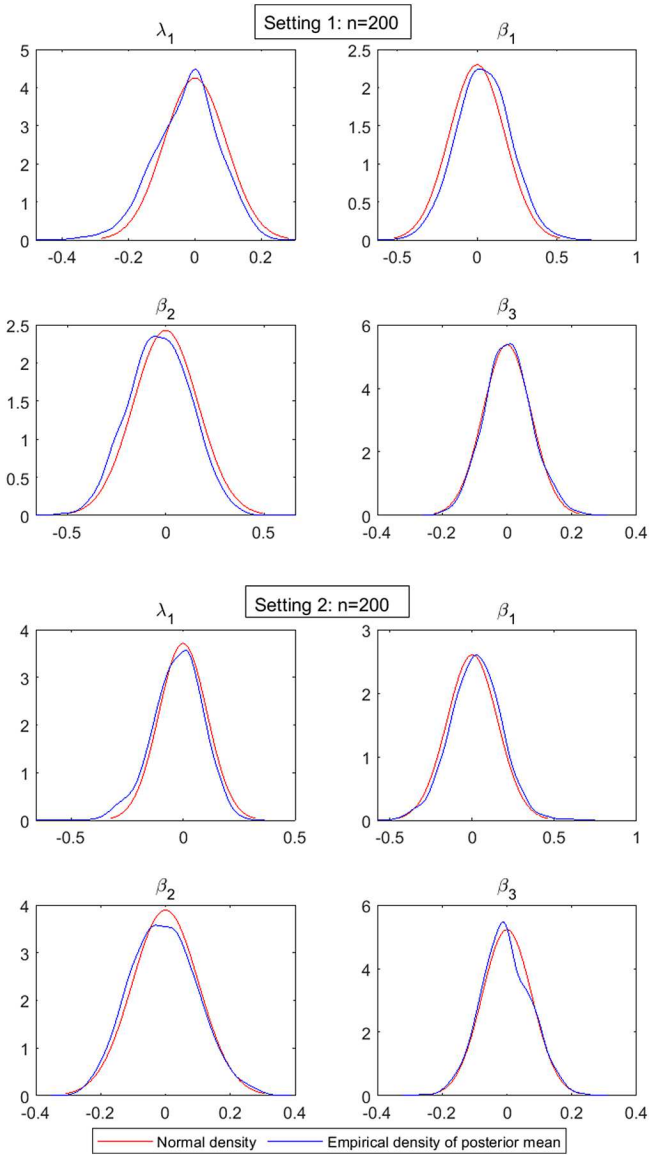


FIGURE 4. Setting 1: x_{2i} i.i.d. Bernoulli(0.5); Setting 2: $X_{2n} = (I_n - 0.2 \times W_{1n})^{-1} \cdot N_n(0, I_n)$. Empirical density of Bayesian posterior mean estimates versus normal density: the SAR Tobit model with $(\lambda_1, \beta_1, \beta_2, \beta_3) = (0.5, -0.5, 1, 1)$.

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

TABLE 3. Estimation results for the SAR Tobit model.

		Setting 1				Setting 2			
		MLE		BE		MLE		BE	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
$n = 200$	λ_1	−0.022	0.100	−0.031	0.100	−0.017	0.111	−0.023	0.110
	β_1	0.030	0.193	0.055	0.190	0.026	0.176	0.025	0.168
	β_2	−0.000	0.163	0.030	0.161	−0.002	0.072	−0.002	0.071
	β_3	−0.002	0.066	−0.004	0.066	−0.001	0.092	−0.005	0.091
	β_4	0.001	0.063	−0.001	0.063	−0.000	0.076	−0.002	0.076
	σ^2	−0.026	0.128	0.016	0.126	−0.045	0.150	0.011	0.143
$n = 400$	λ_1	−0.009	0.067	−0.011	0.066	−0.009	0.063	−0.010	0.061
	β_1	0.007	0.128	0.016	0.121	0.011	0.116	0.007	0.108
	β_2	−0.002	0.123	−0.017	0.121	−0.001	0.055	−0.002	0.054
	β_3	0.000	0.050	−0.001	0.049	0.001	0.050	0.001	0.050
	β_4	0.001	0.051	0.001	0.051	0.002	0.051	0.002	0.051
	σ^2	−0.015	0.089	0.008	0.088	−0.021	0.102	0.006	0.100
$n = 800$	λ_1	−0.004	0.043	−0.006	0.042	−0.005	0.045	−0.004	0.040
	β_1	0.006	0.091	0.011	0.089	0.006	0.091	0.001	0.075
	β_2	−0.001	0.088	−0.008	0.088	−0.001	0.037	0.001	0.037
	β_3	−0.002	0.033	−0.002	0.033	0.000	0.038	0.001	0.038
	β_4	−0.000	0.032	−0.001	0.032	−0.002	0.036	−0.002	0.035
	σ^2	−0.009	0.064	0.002	0.063	−0.009	0.070	0.004	0.069

The values reported in this table are calculated from 1,000 repetitions.
In the 1,000 repetitions of each case, the generatedX_n and W_n are fixed across repetitions, but the disturbances are random draws. Abbreviations: BE, Bayesian estimate; MLE, maximum likelihood estimate; RMSE, root mean squared error.
$(\lambda_1, \beta_1, \beta_2, \beta_3, \beta_4, \sigma^2) = (0.5, -0.5, 1, 1, 1, 1)$.
Setting 1: x_{2i} i.i.d. Bernoulli(0.5); Setting 2: $X_{2n} = (I_n - 0.2 \times W_{1n})^{-1} \cdot N_n(0, I_n)$.

with two spatial lags in both the main equation and the error term, whereas it only takes the exchange algorithm within 4.3 min to finish estimation.

4.3.2. *The SAR Tobit model.* Figure 3 gives the marginal posterior densities of λ_1 and β_k for $k = 1, 2, 3$ and their corresponding normal densities, for the SAR Tobit model with $n = 200$. Plots of posterior densities behave similarly to normal densities. The fit between the two densities continues to improve as n enlarges.[27] Figure 4 summarizes empirical densities of posterior mean estimates of λ_1 and β_k for $k = 1, 2, 3$ and corresponding normal densities. These empirical densities look similar to normal densities.

[27] See Section H of the supplementary appendices for more figures.

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

Table 3 reports biases and RMSEs of MLEs and BEs for the SAR Tobit model. We find that biases and RMSEs of MLEs and BEs are very similar under different scenarios, for all sample sizes.

5. CONCLUSION

In this paper, we apply and extend the theory in CH (2003) to study large sample properties of a posterior pdf and a BE (or QBE) for some spatial econometric models. For a high-order SARAR model, when the sample size is large, Bayesian estimation can have some computational advantages over the classical likelihood based approach. Under some regularity conditions, the posterior pdf of the high-order SARAR model behaves like a normal density function, and the distributions of the posterior mean are asymptotically the same as those of the MLE (or QMLE). These properties also hold for the SAR Tobit model.

Simulation studies show that these normal approximations behave well in most cases even when the sample size is not large. When the sample size is large, such approximations work very well. Our study provides theoretical justifications for the Bayesian approach for spatial models in the classical statistical framework.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0266466620000286

REFERENCES

Andrews, D. (1992) Generic uniform convergence. *Econometric Theory* 8, 241–257.

Chen, C.F. (1985) On asymptotic normality of limiting density functions with Bayesian implications. *Journal of the Royal Statistical Society: Series B* 47(3), 540–546.

Cheng, T., J. Gao, & P.C.B. Phillips (2018) A frequency approach to Bayesian asymptotics. *Journal of Econometrics* 206, 359–378.

Congdon, P. (2006) *Bayesian Statistical Modelling, Second Edition,* John Wiley & Sons, Ltd.

Chernozhukov, V. & H. Hong (2003) An MCMC approach to classical estimation. *Journal of Econometrics* 115(2), 293–346.

Creel, M., J. Gao, H. Hong, & D. Kristensen (2016) Bayesian Indirect Inference and the ABC of GMM. Universitat Autonoma de Barcelona, Monash University, Stanford University, University College London. Working Paper.

Gelman, A., G. Roberts, & W. Gilks (1996) Efficient metropolis jumping rules. *Bayesian Statistics* 5, 599–607.

Golub, G., M. Heath, & G. Wahba (1979) Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* 21, 215–223.

Gupta, A. (2019) Estimation of spatial autoregressions with stochastic weight matrices. *Econometric Theory* 35(2), 417–463.

Gupta, A. & P. Robinson (2015) Inference on higher-order spatial autoregressive models with increasingly many parameters. *Journal of Econometrics* 186, 19–31.

Gupta, A. & P. Robinson (2018) Pseudo maximum likelihood estimation of spatial autoregressive models with increasing dimension. *Journal of Econometrics* 202, 92–107.

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

Haario, H., E. Saksman, & J. Tamminen (2001) An adaptive metropolis algorithm. *Bernoulli* 7, 223–242.

Han, X., C.S. Hsieh, & L. Lee (2017) Estimation and model selection of higher-order spatial autoregressive model: An efficient Bayesian approach. *Regional Science and Urban Economics* 63, 97–120.

Han, X. & L. Lee (2013) Bayesian estimation and model selection for spatial Durbin error model with finite distributed lags. *Regional Science and Urban Economics* 643, 816–837.

Han, X. & L. Lee (2016). Bayesian analysis of spatial panel autoregressive models with time-varying endogenous spatial weight matrices, common factors, and random coefficients. *Journal of Business & Economic Statistics* 34(4), 642–660.

Hoerl, A. & R. Kennard (1970) Ridge regression: Applications to nonorthogonal problems. *Technometrics* 12, 55–68.

Jenish, N. & I.R. Prucha (2009) Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics* 150, 86–98.

Jenish, N. & I.R. Prucha (2012) On spatial processes and asymptotic inference under near-epoch dependence. *Journal of Econometrics* 170, 178–190.

Jin, F. & L. Lee (2013) Cox-type tests for competing spatial autoregressive models with spatial autoregressive disturbances. *Regional Science and Urban Economics* 43, 590–616.

Kelejian, H.H. & I.R. Prucha (1998) A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics* 17, 99–121.

Kim, J.Y. (1998) Large sample properties of posterior densities, Bayesian information criterion and the likelihood principle in nonstationary time series models. *Econometrica* 66(2), 359–380.

Kyung, M., J. Gill, M. Ghosh, & G. Casella (2010) Penalized regression, standard errors and Bayesian lasso. *Bayesian Analysis* 5, 369–412.

Lee, L. (2004) Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72, 1899–1925.

Lee, L. (2007) GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics* 137, 489–514.

Lee, L. & X. Liu (2010) An efficient GMM estimator of spatial autoregressive models with autoregressive disturbances. *Econometric Theory* 26(1), 187–230.

Lee, L. & J. Yu (2016) Identification of spatial Durbin panel models. *Journal of Applied Econometrics* 31(1), 133–162.

Lehmann, E.L. & G. Casella (1998) *Theory of Point Estimation*. Springer Science & Business Media.

LeSage, J. & R.K. Pace (2008) Spatial econometric modeling of origin-destination flows. *Journal of Regional Science* 48, 941–967.

LeSage, J.P. & R.K. Pace (2009) *Introduction to Spatial Econometrics*. CRC Press.

Li, Q. & N. Lin (2010) The Bayesian elastic net. *Bayesian Analysis* 5, 151–170.

Li, R.C. (1998) Relative perturbation theory: I. Eigenvalue and singular value variations. *SIAM Journal on Matrix Analysis and Applications* 19(4), 956–982.

Marquardt, D. (1970) Generalized inverses, ridge regression, bias linear estimation and nonlinear estimation. *Technometrics* 12, 591–612.

Murray, I., Z. Ghahramani, & D. MacKay (2006) MCMC for doubly-intractable distributions. arXiv preprint, arXiv:1206.6848.

Ord, K. (1975) Estimation methods for models of spatial interactions. *Journal of American Statistical Association* 70, 120–126.

Pace, R.K., J. LeSage, & S. Zhu (2012) Spatial dependence in regressors and its effect on performance of likelihood-based and instrumental variable estimator. *Advances in Econometrics* 30, 257–295.

Park, T. & G. Casella (2008) The Bayesian lasso. *Journal of the American Statistical Association* 103, 681–686.

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

Phillips, P.C.B. & W. Ploberger (1996) An asymptotic theory of Bayesian inference for time series. *Econometrica* 64(2), 381–412.

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

Shiller, R. (1973) A distributed lag estimator derived from smoothness priors. *Econometrica* 41(4), 775–788.
Qu, X. & L. Lee (2013) Locally most powerful tests for spatial interactions in the simultaneous sar Tobit model. *Regional Science and Urban Economics* 43, 307–321.
Qu, X. & L. Lee (2015) Estimating a spatial autoregressive model with an endogenous spatial weight matrix. Journal of Econometrics 184, 209–232.
Raiffa, H. & R. Schlaifer (1961) *Applied Statistical Design Theory*. Clinton Press.
Roberts, G. & J.S. Rosenthal (2009) Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics* 18, 349–367.
Smith, T.E. (2009) Estimation bias in spatial models with strongly connected weight matrices. *Geographical Analysis* 41, 307–332.
Theil, H. & A. Goldberger (1961) On Pure and Mixed Statistical Estimation in Economics. *International Economic Review,* 2(1), 65–78.
Theil, H. (1963) On the use of incomplete prior information in regression analysis. *Journal of the American Statistical Association* 58, 401–414.
Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B* 58, 267–288.
Van der Vaart, A.W. (1998) *Asymptotic Statistics*. Cambridge University Press.
Van Erp, S., D.L. Oberski, & J. Mulder (2019) Shrinkage priors for Bayesian penalized regression. *Journal of Mathematical Psychology* 89, 31–50.
Xu, X. & L. Lee (2015) Maximum likelihood estimation of a spatial autoregressive Tobit model. *Journal of Econometrics* 188, 264–280.

A. SOME DISCUSSION ABOUT ASSUMPTION 3.5

Without the compactness of the parameter space, we need Assumption 3.5 to make sure that the QMLE $\hat{\lambda}$ and $\hat{\rho}$ will not converge to the boundary of the parameter space. Assumption 3.5(i) for our model is essentially Assump. 8 in Lee (2004) for the SAR model and part of Assump. 7(i) in Jin and Lee (2013) for the SARAR(1,1) model. In Assumption 3.5(ii), because it is possible that $R_n(\rho)$ is singular on the boundary, we let $\eta \in (0,1)$. Since $\sup_{n,\,\rho\in\Theta_\rho} ||M_n(\rho)||_\infty < 1$, we have $\sup_{n,\,\rho\in\overline{\eta\Theta_\rho}} ||M_n(\rho)||_\infty < \eta$. As a result,

$$\sup_{n,\,\rho\in\overline{\eta\Theta_\rho}} ||R_n(\rho)^{-1}||_\infty = \sup_{n,\,\rho\in\overline{\eta\Theta_\rho}} ||\sum_{l=0}^{\infty} M_n(\rho)^l||_\infty \le \frac{1}{1-\eta} < \infty. \tag{A.1}$$

With this result, Assumption 3.5(ii) implies that

$$\inf_{n,\,\rho\in\overline{\eta\Theta_\rho}} \min \text{eig}[R_n(\rho)'R_n(\rho)] = \inf_{n,\,\rho\in\overline{\eta\Theta_\rho}} \left\{\max \text{eig}[R_n(\rho)'R_n(\rho)]^{-1}\right\}^{-1}$$
$$\geqslant \inf_{n,\,\rho\in\overline{\eta\Theta_\rho}} \left\{\left\|[R_n(\rho)'R_n(\rho)]^{-1}\right\|_1\right\}^{-1} \geqslant \inf_{n,\,\rho\in\overline{\eta\Theta_\rho}} \left\{\left\|R_n(\rho)^{-1}\right\|_1 \left\|R_n(\rho)^{-1}\right\|_\infty\right\}^{-1} > 0. \tag{A.2}$$

Assumption 3.5(iii) is related to the identification of ρ and σ^2. The function $f(x) \equiv x - \ln x - 1$ is strictly decreasing on $(0,1)$ and strictly increasing on $(1,\infty)$. Thus, it has a unique minimum value $f(1) = 0$. As a result, for any positive definite $n \times n$ matrix A,

$$f(A) \equiv \frac{1}{n}[tr(A) - \ln(A) - n] \geqslant 0, \tag{A.3}$$

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

and the equality holds only when all eigenvalues of A equal 1, which implies $A = I$. A sufficient condition for $f(R_n^{'-1}R_n(\rho)'R_n(\rho)R_n^{-1}) > 0$ is that the q^2+q+1 matrices, $\{M'_{kn}M_{jn}\}_{j,k=1}^{q}$, $\{M'_{kn}+M_{kn}\}_{k=1}^{q}$ and I_n, are linearly independent. This condition will be satisfied in most empirical applications. Under this sufficient condition, when $\rho \neq \rho_0$, we can show by a contradiction argument that $R_n^{'-1}R'_n(\rho)R_n(\rho)R_n^{-1} \neq I_n$. Thus, $f(R_n^{'-1}R_n(\rho)'R_n(\rho)R_n^{-1}) > 0$. A similar linearly independent condition on matrices is also exploited in Lee and Yu (2016) for spatial panel data models.

To deal with the limiting case as $n \to \infty$, we establish the following lemma. It implies that if the sample variance of eigenvalues of a sequence of positive definite matrices A_n is uniformly bounded away from 0, which means that most eigenvalues are not too close to each other, $f(A_n)$ is also uniformly bounded away from 0. That is why Assumption 3.5(iii) is about the sample variances of some eigenvalues.

LEMMA A.1. *A_n is an $n \times n$ positive definite matrix satisfying $\frac{1}{n}\sum_{i=1}^{n}(eig_i(A_n)-1)^2 \geqslant C^2$ for some constant $C > 0$ and for all n, where $\mathrm{eig_i(A_n)} \leqslant \mathrm{D}$ for a constant $D > 1 + \frac{C}{2}$ for all n and i. Then there is a constant $\delta(C,D) > 0$ such that $f(A_n) \equiv \frac{1}{n}[tr(A_n) - \ln(A_n) - n] \geqslant \delta(C,D)$ for all n.*

Proof of Lemma A.1: Denote $I = \{i \in \{1,2,\ldots,n\} : |eig_i(A_n) - 1| > \frac{C}{2}\}$. The first condition implies that $I \neq \emptyset$. Then

$$C^2 \leqslant \frac{1}{n}\sum_{i=1}^{n}(\mathrm{eig_i(A_n)}-1)^2 \leqslant \frac{1}{\mathrm{n}}\left[(\mathrm{D}-1)^2|\mathrm{I}| + (\mathrm{n}-|\mathrm{I}|)\left(\frac{\mathrm{C}}{2}\right)^2\right]. \tag{A.4}$$

Solving this inequality, we have

$$\frac{|I|}{n} \geqslant \frac{3C^2}{4(D-1)^2 - C^2}. \tag{A.5}$$

Denote $\Delta(C) \equiv \inf_{x>0:|x-1|>C/2}(x - \ln x - 1) > 0$. Thus,

$$\frac{1}{n}[\mathrm{tr}(A_n) - \ln(A_n) - n] \geqslant \frac{1}{n}\sum_{i\in I}\left[\mathrm{eig_i(A_n)} - \ln \mathrm{eig_i(A_n)} - 1\right]$$

$$\geqslant \frac{|I|}{n}\Delta(C) = \frac{3C^2\Delta(C)}{4(D-1)^2 - C^2} \equiv \delta(C,D) > 0. \tag{A.6}$$

□

B. PROOFS

Since the proof of Theorem 1 is based on Thm. 1 of (CH (2003)), we first cite their result here for reference. A general extremum estimator is defined to be $\hat{\theta} = \arg\max_{\theta\in\Theta}\ln L_n(\theta)$. Given a prior $\pi(\theta)$ with its support on Θ, define the posterior quasi pdf as

$$p_n(\theta) = \frac{L_n(\theta)\pi(\theta)}{\int_\Theta L_n(\phi)\pi(\phi)d\phi}. \tag{B.1}$$

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

LEMMA B.2. *(Thm. 1, CH, 2003) Suppose the following conditions hold: (1)* $\Theta \subseteq \mathbb{R}^K$ *is convex and* $\theta_0 \in \Theta^0$, *the interior of* Θ. *(2) The prior pdf* $\pi(\theta)$ *is continuous at* θ_0, *bounded,*[28] *and* $\int_\Theta ||\theta||_2^\kappa \pi(\theta) d\theta < \infty$ *for some constant* $\kappa \geqslant 0$.[29] *(3) For any* $\delta > 0$, *there exists an* $\epsilon > 0$, *such that*

$$\liminf_{n\to\infty} \Pr\left\{ \sup_{\theta:||\theta-\theta_0||_2>\delta} \frac{1}{n}\left[\ln L_n(\theta) - \ln L_n(\theta_0)\right] \leqslant -\epsilon \right\} = 1. \tag{B.2}$$

(4) For any θ *in an open neighborhood of* θ_0, *(i)* $\ln L_n(\theta) - \ln L_n(\theta_0) = (\theta-\theta_0)'\Delta_n - \frac{1}{2}(\theta-\theta_0)'(nJ_n)(\theta-\theta_0) + R_n(\theta)$, *(ii)* $\Omega_n^{-1/2}\Delta_n/\sqrt{n} \xrightarrow{d} N(0,I)$, *(iii) the smallest eigenvalues of* $J_n = O(1)$ *and* $\Omega_n = O(1)$ *are uniformly bounded away from 0, (iv) for each* $\epsilon > 0$, *there is a sufficiently small* $\delta > 0$ *and large* $M > 0$ *such that* $\limsup_{n\to\infty} \Pr\left\{\sup_{\theta:M/\sqrt{n}\leqslant||\theta-\theta_0||_2\leqslant\delta} \frac{|R_n(\theta)|}{n||\theta-\theta_0||_2^2} > \epsilon\right\} < \epsilon$ *and* $\limsup_{n\to\infty} \Pr\left\{\sup_{\theta:||\theta-\theta_0||_2\leqslant M/\sqrt{n}} |R_n(\theta)| > \epsilon\right\} = 0$.

Let $h \equiv \sqrt{n}(\theta-\theta_0) - J_n^{-1}\frac{1}{\sqrt{n}}\Delta_n$ *and define the quasi pdf of* h *as* $p_n^*(h) = \frac{1}{\sqrt{n}} p_n\left(\frac{h}{\sqrt{n}} + \theta_0 + J_n^{-1}\frac{1}{n}\Delta_n\right)$. *Then* $\left\|p_n^*(h) - \bar{p}_n(h)\right\|_{TVM(\kappa)} \xrightarrow{p} 0$, *where* $\bar{p}_n(h) = (2\pi)^{-K/2}|J_n|^{1/2}\exp\left(-\frac{1}{2}h'J_nh\right)$ *is the pdf of* $N(0, J_n^{-1})$.

LEMMA B.3. *(Lem. 1, CH, 2003) In Lemma B.2, condition (3) holds if there is a nonstochastic continuous function* $M_n(\theta)$ *such that (i) for any* $\delta > 0$, $\limsup_{n\to\infty}[\sup_{\theta\in\Theta:||\theta-\theta_0||_2>\delta} M_n(\theta) - M_n(\theta_0)] < 0$, *and (ii)* $\sup_{\theta\in\Theta}|\frac{1}{n}\ln L_n(\theta) - M_n(\theta)| = o_p(1)$.

LEMMA B.4. *(Lem. 2, CH, 2003) In Lemma B.2, given conditions (1) and (3), condition (4) holds with* $\Delta_n = \frac{\partial \ln L_n(\theta_0)}{\partial\theta}$ *and* $J_n = -\frac{\partial^2 M_n(\theta_0)}{\partial\theta\partial\theta'} = O(1)$, *if*

(i) $\ln L_n(\theta)$ *and* $M_n(\theta)$ *are twice continuously differentiable in a neighborhood of* θ_0,

(ii) *there is an* $\Omega_n = O(1)$ *such that* $\Omega_n^{-1/2} n^{-1/2}\frac{\partial \ln L_n(\theta_0)}{\partial\theta} \xrightarrow{d} N(0, I_n)$, J_n *and* Ω_n *are uniformly positive definite, and*

(iii) *for some* $\delta > 0$ *and for each* $\epsilon > 0$,

$$\limsup_{n\to\infty} \Pr\left\{ \sup_{\theta\in\Theta:||\theta-\theta_0||_2<\delta} \left| \frac{1}{n}\frac{\partial^2 \ln L_n(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 M_n(\theta)}{\partial\theta\partial\theta'} \right| > \epsilon \right\} = 0.$$

Proof of Lemma 2: (1) First, consider σ^2. By Assumptions 3.3 and 3.4(i),

$$\begin{aligned} &\sup_{n,\rho} \max \operatorname{eig}\left[R_n(\rho)'R_n(\rho)\right] \leqslant \sup_{n,\rho} \left\|R_n(\rho)'R_n(\rho)\right\|_1 \\ &\leqslant \sup_{n,\rho} \left\|R_n(\rho)'\right\|_1 \|R_n(\rho)\|_1 = \sup_{n,\rho} \|R_n(\rho)\|_\infty \|R_n(\rho)\|_1 < C_1 < \infty \end{aligned} \tag{B.3}$$

[28] From the proof of Thm. in CH (2003), we see that the continuity of $\pi(\theta)$ at θ_0 and the boundedness of $\pi(\cdot)$ are sufficient to replace the continuity of $\pi(\theta)$ on Θ.

[29] From CH (2003), the compactness of Θ can be dropped if we add the condition $\int \pi(\theta)||\theta||_2^\kappa d\theta < \infty$.

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

for some constant C_1, where the first inequality holds because eigenvalues of a matrix are bounded by any of its norms (the spectral radius theorem). Thus,

$$\hat{\sigma}_n^2(\lambda,\rho)=\frac{1}{n}\left[M_{R_n(\rho)X_n}S_n(\lambda)Y_n\right]'\left[R_n(\rho)'R_n(\rho)\right]\left[M_{R_n(\rho)X_n}S_n(\lambda)Y_n\right]$$
$$\leqslant C_1\frac{1}{n}[S_n(\lambda)Y_n]'M_{R_n(\rho)X_n}[S_n(\lambda)Y_n]\leqslant C_1\frac{1}{n}[S_n(\lambda)Y_n]'[S_n(\lambda)Y_n]. \tag{B.4}$$

By Assumption 3.4(i), Lemma 1, and Minkowski's inequality, $C_2\equiv\sup_{i,n,\lambda}||S_n(\lambda)_{i\cdot}Y_n||_{L^{4+\Delta}}<\infty$. As a result,

$$\sup_{n,\lambda,\rho}E\hat{\sigma}_n^2(\lambda,\rho)\leqslant\sup_{n,\lambda}C_1\frac{1}{n}\sum_{i=1}^{n}E[S_n(\lambda)_{i\cdot}Y_n]^2\leqslant C_1C_2^2<\infty. \tag{B.5}$$

That is to say, $\hat{\sigma}_n^2(\lambda,\rho)=O_p(1)$ uniformly in λ and ρ. Hence, from $\frac{1}{n}\frac{\partial\ln L_n(\theta)}{\partial\sigma^2}=-\frac{1}{2\sigma^2}+\frac{1}{2\sigma^4}\hat{\sigma}_n^2(\lambda,\rho)$, there exists a constant C_3 such that when $\sigma^2\geqslant C_3$, $\frac{1}{n}\frac{\partial\ln L_n(\theta)}{\partial\sigma^2}<0$ with probability approaching 1 (WPA1). On the other hand, by Assumption 3.4(ii), when $\sigma^2<L_\sigma$, $\frac{1}{n}\frac{\partial\ln L_n(\theta)}{\partial\sigma^2}>0$ WPA1. Denote $\widetilde{\Theta}_{\sigma^2}\equiv[L_\sigma,C_3]$. As a result, $\sup_{\theta\in\Theta-B(\theta_0,\gamma_0)}\ln L_n(\theta)=\sup_{\theta\in\Theta-B(\theta_0,\gamma_0):\sigma^2\in\widetilde{\Theta}_{\sigma^2}}\ln L_n(\theta)$ WPA1.

(2) Second, consider β. By Assumption 3.4(iii), $[\frac{1}{n}(R_n(\rho)X_n)'(R_n(\rho)X_n)]^{-1}=O_p(1)$ uniformly in ρ. Thus, for some constant $C_4>0$ not depending on ρ,

$$\hat{\beta}_n(\lambda,\rho)'\hat{\beta}_n(\lambda,\rho)$$
$$=[R_n(\rho)S_n(\lambda)Y_n]'(R_n(\rho)X_n)[(R_n(\rho)X_n)'(R_n(\rho)X_n)]^{-2}(R_n(\rho)X_n)'R_n(\rho)S_n(\lambda)Y_n$$
$$\leqslant C_4[R_n(\rho)S_n(\lambda)Y_n]'(R_n(\rho)X_n)[(R_n(\rho)X_n)'(R_n(\rho)X_n)]^{-1}(R_n(\rho)X_n)'R_n(\rho)S_n(\lambda)Y_n+o_p(1)$$
$$=C_4\left\|P_{R_n(\rho)X_n}R_n(\rho)S_n(\lambda)Y_n\right\|_2^2+o_p(1)\leqslant C_4\left\|R_n(\rho)S_n(\lambda)Y_n\right\|_2^2+o_p(1), \tag{B.6}$$

where $P_{R_n(\rho)X_n}\equiv(R_n(\rho)X_n)[(R_n(\rho)X_n)'(R_n(\rho)X_n)]^{-1}(R_n(\rho)X_n)'$ is a projection matrix and the $o_p(1)$ is uniformly in ρ. Again, by Assumption 3.4(i), Lemma 1, and Minkowski's inequality, we have

$$\sup_{i,n,\rho,\lambda}||R_n(\rho)_{i\cdot}S_n(\lambda)Y_n||_{L^{4+\Delta}}<\infty.$$

Thus, $\hat{\beta}_n(\lambda,\rho)=O_p(1)$ uniformly in λ and ρ. That is to say, there is a large enough compact set $\widetilde{\Theta}_\beta$, such that WPA1, $\hat{\beta}_n(\lambda,\rho)\in\widetilde{\Theta}_\beta$ holds for all (λ,ρ) in the parameter space.

When $\sigma^2\in\widetilde{\Theta}_{\sigma^2}$, because

$$\frac{1}{n}\frac{\partial^2\ln L_n(\theta)}{\partial\beta\partial\beta'}=-\frac{X_n'R_n(\rho)'R_n(\rho)X_n}{n\sigma^2}\leqslant-\frac{X_n'R_n(\rho)'R_n(\rho)X_n}{nC_3}, \tag{B.7}$$

$\frac{1}{n}\ln L_n(\theta)$ is a strictly concave function with respect to β. Therefore, for any (λ,ρ), $\beta\notin\widetilde{\Theta}_\beta$, and $\sigma^2\in\widetilde{\Theta}_{\sigma^2}$, the function $g(t)\equiv\frac{1}{n}\ln L_n(\lambda,\hat{\beta}_n(\lambda,\rho)+t[\beta-\hat{\beta}_n(\lambda,\rho)],\rho,\sigma^2)$ is a strictly concave function, which is maximized at $t=0$. Thus, $g(t)$ is a strictly decreasing function on $[0,1]$. Thus, WPA1,

$$\sup_{\theta\in\Theta-B(\theta_0,\gamma_0)}\ln L_n(\theta)=\sup_{\theta\in\Theta-B(\theta_0,\gamma_0):\sigma^2\in\widetilde{\Theta}_{\sigma^2},\beta\in\widetilde{\Theta}_\beta}\ln L_n(\theta). \tag{B.8}$$

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

Then this lemma holds with $\widetilde{\Theta} \equiv \Theta_\lambda \times \widetilde{\Theta}_\beta \times \Theta_\rho \times \widetilde{\Theta}_{\sigma^2}$. □

Proof of Theorem 1: We will verify the conditions of Lemma B.2 to establish this theorem.

(1) We shall utilize Lemma B.3 to verify condition (3) of Lemma B.2, the identification uniqueness condition, in three steps. With Lemma 2, it suffices to show that for any $\delta > 0$, there exists an $\epsilon > 0$, such that

$$\liminf_{n\to\infty} \Pr\left\{ \sup_{\theta\in\widetilde{\Theta}:||\theta-\theta_0||_2>\delta} \frac{1}{n}\left[\ln L_n(\theta) - \ln L_n(\theta_0)\right] \leqslant -\epsilon \right\} = 1. \tag{B.9}$$

For any $\eta \in (0,1)$, denote $\widetilde{\Theta}_\eta \equiv \overline{\eta\Theta_\lambda} \times \widetilde{\Theta}_\beta \times \overline{\eta\Theta_\rho} \times \widetilde{\Theta}_{\sigma^2}$, which is a compact set. We will only consider those η large enough such that $\theta_0 \in \widetilde{\Theta}_\eta$.

(i) We aim to show that $\{\frac{1}{n}E\ln L_n(\theta)\}$ is equicontinuous on $\widetilde{\Theta}_\eta$, and it is sufficient to prove that its derivatives are uniformly bounded in n on $\widetilde{\Theta}_\eta$.
Note that

$$\frac{1}{n}\frac{\partial \ln L_n(\theta)}{\partial \lambda_j} = -\frac{1}{n}tr[S_n(\lambda)^{-1}W_{jn}] + \frac{1}{n\sigma^2}\left\{(W_{jn}Y_n)'R_n(\rho)'R_n(\rho)\left[S_n(\lambda)Y_n - X_n\beta\right]\right\}. \tag{B.10}$$

Because $\sup_{n,\lambda\in\overline{\eta\Theta_\lambda}} \|W_n(\lambda)\|_\infty \leqslant \eta < 1$ by equation (3.3),

$$\begin{aligned}
&\sup_{n,\lambda\in\overline{\eta\Theta_\lambda}} ||S_n(\lambda)^{-1}W_{jn}||_\infty = \sup_{n,\lambda\in\overline{\eta\Theta_\lambda}} \left\| \sum_{l=0}^{\infty} W_n^l(\lambda)W_{jn} \right\|_\infty \\
&\leqslant \sup_{n,\lambda\in\overline{\eta\Theta_\lambda}} \sum_{l=0}^{\infty} \|W_n(\lambda)\|_\infty^l \left\|W_{jn}\right\|_\infty \leqslant \frac{\sup_n \left\|W_{jn}\right\|_\infty}{1-\eta} < \infty.
\end{aligned} \tag{B.11}$$

As a consequence, $\sup_{n,\lambda\in\overline{\eta\Theta_\lambda}} \left|\frac{1}{n}tr[S_n(\lambda)^{-1}W_{jn}]\right| < \infty$. By Assumption 3.3 and equation (A.1),

$$\begin{aligned}
C_1 &\equiv \sup_{n,\rho\in\Theta_\rho} \max \operatorname{eig}\left[R_n(\rho)'R_n(\rho)\right] \\
&\equiv \sup_{n,\rho\in\Theta_\rho} ||R_n(\rho)||_2^2 \leqslant \sup_{n,\rho\in\Theta_\rho} ||R_n(\rho)||_1 ||R_n(\rho)||_\infty < \infty.
\end{aligned} \tag{B.12}$$

Then by the inequality $|x'Ay| \leqslant \max|eig(A)| \cdot (x'x)^{1/2}(y'y)^{1/2}$ (A is a symmetric matrix), the Cauchy–Schwarz inequality, Lemma 1 and Assumption 3.2, we have

$$\begin{aligned}
&\sup_n E \sup_{\theta\in\widetilde{\Theta}} \frac{1}{n}\left|\left\{(W_{jn}Y_n)'R_n(\rho)'R_n(\rho)\left[S_n(\lambda)Y_n - X_n\beta\right]\right\}\right| \\
\leqslant& C_1 \sup_n E \sup_{\theta\in\widetilde{\Theta}} \frac{1}{n}\left\{(W_{jn}Y_n)'(W_{jn}Y_n)\left[S_n(\lambda)Y_n - X_n\beta\right]'\left[S_n(\lambda)Y_n - X_n\beta\right]\right\}^{1/2} \\
\leqslant& C_1 \left\{ \sup_n \frac{1}{n} E\left[(W_{jn}Y_n)'(W_{jn}Y_n)\right] \cdot \sup_n E \sup_{\theta\in\widetilde{\Theta}} \frac{1}{n}\left[S_n(\lambda)Y_n - X_n\beta\right]'\left[S_n(\lambda)Y_n - X_n\beta\right] \right\}^{1/2} \\
<& \infty.
\end{aligned} \tag{B.13}$$

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

So, $\sup_n E\sup_{\theta\in\widetilde{\Theta}}\left|\frac{1}{n}\frac{\partial \ln L_n(\theta)}{\partial\lambda}\right|<\infty$.

Note that $\frac{\partial \ln L_n(\theta)}{\partial\beta_l}=\frac{1}{\sigma^2}\{X'_{\cdot l,n}R_n(\rho)'R_n(\rho)[S_n(\lambda)Y_n-X_n\beta]\}$, where $X_{\cdot l,n}$ is the lth column of X_n. Again, by the inequality $|x'Ay|\leqslant \max|eig(A)|\cdot(x'x)^{1/2}(y'y)^{1/2}$ (A is a symmetric matrix) and the Cauchy–Schwarz inequality,

$$\begin{aligned}
&\sup_n \mathrm{E}\sup_{\theta\in\widetilde{\Theta}}\frac{1}{n}\left|\left\{X'_{\cdot l,n}R_n(\rho)'R_n(\rho)\,[S_n(\lambda)Y_n-X_n\beta]\right\}\right|\\
\leqslant& C_1\sup_n E\sup_{\theta\in\widetilde{\Theta}}\frac{1}{n}\left\{X'_{\cdot l,n}X_{\cdot l,n}\,[S_n(\lambda)Y_n-X_n\beta]'\,[S_n(\lambda)Y_n-X_n\beta]\right\}^{1/2}\\
\leqslant& C_1\left\{\sup_n\frac{1}{n}\left(X'_{\cdot l,n}X_{\cdot l,n}\right)\cdot\sup_n E\sup_{\theta\in\widetilde{\Theta}}\frac{1}{n}\,[S_n(\lambda)Y_n-X_n\beta]'\,[S_n(\lambda)Y_n-X_n\beta]\right\}^{1/2}<\infty.
\end{aligned}\tag{B.14}$$

So, $\sup_n \mathrm{E}\sup_{\theta\in\widetilde{\Theta}_\eta}\left|\frac{1}{n}\frac{\partial \ln L_n(\theta)}{\partial\beta_l}\right|<\infty$.

For $\frac{\partial \ln L_n(\theta)}{\partial\sigma^2}=[S_n(\lambda)Y_n-X_n\beta]'R_n(\rho)'R_n(\rho)\,[S_n(\lambda)Y_n-X_n\beta]/2\sigma^4$, similarly to equation (2.14), we have $\sup_n \mathrm{E}\sup_{\theta\in\widetilde{\Theta}}\left|\frac{1}{n}\frac{\partial \ln L_n(\theta)}{\partial\sigma^2}\right|<\infty$.

For

$$\begin{aligned}
\frac{\partial \ln L_n(\theta)}{\partial\rho_k}=&\frac{[S_n(\lambda)Y_n-X_n\beta]'\left[M'_{kn}R_n(\rho)+R_n(\rho)'M_{kn}\right][S_n(\lambda)Y_n-X_n\beta]}{2\sigma^2}\\
&-\frac{1}{n}tr[R_n(\rho)^{-1}M_{kn}],
\end{aligned}\tag{B.15}$$

similarly to $\sup_{n,\lambda\in\overline{\eta\Theta_\lambda}}\left|\frac{1}{n}tr[S_n(\lambda)^{-1}W_{jn}]\right|<\infty$, we have $\sup_{n,\rho\in\overline{\eta\Theta_\rho}}\left|\frac{1}{n}tr[S_n(\lambda)^{-1}W_{jn}]\right|<\infty$. Because $\sup_{n,\rho}||M'_{kn}R_n(\rho)+R_n(\rho)'M_{kn}||_\infty<\infty$, $\sup_n E\sup_{\theta\in\widetilde{\Theta}_\eta}\left|\frac{1}{n}\frac{\partial \ln L_n(\theta)}{\partial\rho_k}\right|<\infty$.

(ii) In this and the next steps, we shall show that θ_0 is identifiable. It follows from $S_n(\lambda)Y_n-X_n\beta=[S_n(\lambda)S_n^{-1}X_n\beta_0-X_n\beta]+S_n(\lambda)S_n^{-1}R_n^{-1}V_n$ that

$$\begin{aligned}
E\ln L_n(\theta)=&-\frac{n}{2}\ln 2\pi\sigma^2+\ln|S_n(\lambda)R_n(\rho)|\\
&-\frac{\sigma_0^2}{2\sigma^2}tr\left\{[S_n(\lambda)S_n^{-1}R_n^{-1}]'R_n(\rho)'R_n(\rho)[S_n(\lambda)S_n^{-1}R_n^{-1}]\right\}\\
&-\frac{1}{2\sigma^2}E\left\|R_n(\rho)[S_n(\lambda)S_n^{-1}X_n\beta_0-X_n\beta]\right\|_2^2.
\end{aligned}\tag{B.16}$$

And $E\ln L_n(\theta_0)=-\frac{n}{2}\ln 2\pi\sigma_0^2+\ln|S_nR_n|-\frac{n}{2}$. As a result, when $\theta\in\Theta$,

$$\begin{aligned}
&\frac{1}{n}E\ln L_n(\theta_0)-\frac{1}{n}E\ln L_n(\theta)\\
=&-\frac{1}{2}\ln\frac{\sigma_0^2}{\sigma^2}+\frac{1}{n}\ln|S_nR_nS_n^{-1}(\lambda)R_n^{-1}(\rho)|\\
&-\frac{1}{2}+\frac{\sigma_0^2}{2n\sigma^2}tr\left\{[S_n(\lambda)S_n^{-1}R_n^{-1}]'R_n(\rho)'R_n(\rho)[S_n(\lambda)S_n^{-1}R_n^{-1}]\right\}
\end{aligned}$$

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

$$+\frac{1}{2n\sigma^2}E\left\|R_n(\rho)[S_n(\lambda)S_n^{-1}X_n\beta_0-X_n\beta]\right\|_2^2$$
$$=\frac{1}{2}f\left(\frac{\sigma_0^2}{\sigma^2}[R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}]'[R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}]\right)$$
$$+\frac{1}{2n\sigma^2}E\left\|R_n(\rho)[S_n(\lambda)S_n^{-1}X_n\beta_0-X_n\beta]\right\|_2^2\geqslant 0, \tag{B.17}$$

where $f(A)\equiv\frac{1}{n}[tr(A)-\ln(A)-n]$ for any positive definite $n\times n$ matrix A.

(iii) Next, we aim to verify condition (i) in Lemma B.3: for any $\gamma\in(0,\gamma_0)$, where γ_0 is defined in Lemma 2,

$$\liminf_{n\to\infty}\inf_{\theta\in\widetilde{\Theta}:||\theta-\theta_0||_2>\gamma}\frac{1}{n}[E\ln L_n(\theta_0)-E\ln L_n(\theta)]>0. \tag{B.18}$$

If it does not hold, there is a $\gamma\in(0,\gamma_0)$ such that

$$\liminf_{n\to\infty}\inf_{\theta\in\widetilde{\Theta}:||\theta-\theta_0||_2>\gamma}\frac{1}{n}[E\ln L_n(\theta_0)-E\ln L_n(\theta)]=0. \tag{B.19}$$

Then there is a sequence $\{\theta_{m(n)}\}\subseteq\widetilde{\Theta}-B(\theta_0,\gamma)$ such that

$$\lim_{n\to\infty}\frac{1}{m(n)}[E\ln L_{m(n)}(\theta_0)-E\ln L_{m(n)}(\theta_{m(n)})]=0. \tag{B.20}$$

Then it follows from equation (1.2) and (2.17) that

$$\frac{1}{m(n)}E\ln L_{m(n)}(\theta_0)-\frac{1}{m(n)}E\ln L_{m(n)}(\theta)$$
$$\geqslant\frac{1}{2m(n)\sigma^2}E\left\|R_{m(n)}(\rho)[S_{m(n)}(\lambda)S_{m(n)}^{-1}X_{m(n)}\beta_0-X_{m(n)}\beta]\right\|_2^2$$
$$\geqslant\frac{C_2}{2m(n)\sigma^2}E\left\|S_{m(n)}(\lambda)S_{m(n)}^{-1}X_{m(n)}\beta_0-X_{m(n)}\beta\right\|_2^2\geqslant 0 \tag{B.21}$$

for some constant $C_2>0$ that depends on neither $\rho\in\eta\Theta_\rho$ nor n. Because

$$S_n(\lambda)S_n^{-1}X_n\beta_0-X_n\beta=X_n(\beta_0-\beta)+\sum_{j=1}^{p}(\lambda_{j0}-\lambda_j)W_nS_n^{-1}X_n\beta_0$$
$$=[X_n,\,W_{1n}S_n^{-1}X_n\beta_0,\ldots,W_{pn}S_n^{-1}X_n\beta_0][(\beta_0-\beta)',\,(\lambda_0-\lambda)']', \tag{B.22}$$

by Assumption 3.5(i) and equation (2.17), $\beta_{m(n)}\to\beta_0$ and $\lambda_{m(n)}\to\lambda_0$. Next, because

$$R_{m(n)}(\rho_{m(n)})S_{m(n)}(\lambda_{m(n)})S_{m(n)}^{-1}R_{m(n)}^{-1}$$
$$=R_{m(n)}(\rho_{m(n)})R_{m(n)}^{-1}+R_{m(n)}(\rho_{m(n)})W_{m(n)}(\lambda_0-\lambda_{m(n)})S_{m(n)}^{-1}R_{m(n)}^{-1}\equiv P_{1n}+P_{2n}, \tag{B.23}$$

we have

$$[R_{m(n)}(\rho_{m(n)})S_{m(n)}(\lambda_{m(n)})S_{m(n)}^{-1}R_{m(n)}^{-1}]'[R_{m(n)}(\rho_{m(n)})S_{m(n)}(\lambda_{m(n)})S_{m(n)}^{-1}R_{m(n)}^{-1}]$$
$$=P_{1n}'P_{1n}+(P_{1n}'P_{2n}+P_{2n}'P_{1n}+P_{2n}'P_{2n}). \tag{B.24}$$

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

Because $\lambda_{m(n)} \to \lambda_0$, by Assumptions 3.3 and the boundedness of Θ_ρ, we have $\sup_{n,\rho} ||R_n(\rho_{m(n)})||_2^2 \leqslant \sup_{n,\rho} ||R_n(\rho_{m(n)})||_1 \cdot ||R_n(\rho_{m(n)})||_\infty < \infty$, $\lim_{n\to\infty} ||W_{m(n)}(\lambda_0 - \lambda_{m(n)})||_2 = 0$, and $\sup_n ||S_{m(n)}||_2 < \infty$. By Assumption 3.5(ii) and equation (1.1), $\sup_n ||R_{m(n)}^{-1}||_2 \leq \sup_n \left[||R_{m(n)}^{-1}||_1 \cdot ||R_{m(n)}^{-1}||_\infty\right]^{1/2} < \infty$. Consequently, $\lim_{n\to\infty} ||P_{2n}||_2 = 0$ and $\sup_{n\to\infty} ||P_{1n}||_2 < \infty$. Hence, $\lim_{n\to\infty} ||P_{1n}' P_{2n} + P_{2n}' P_{1n} + P_{2n}' P_{2n}||_2 = 0$. For any two symmetric matrices A and B, $|eig_i(A) - eig_i(B)| \leqslant ||A - B||_2$, where eig_i is the ith greatest eigenvalue (Li, 1998). Using this conclusion, we have

$$\lim_{n\to\infty} \max_{i=1,\ldots,n} \left|eig_i\left[(P_{1n}+P_{2n})'(P_{1n}+P_{2n})\right] - eig_i\left(P_{1n}'P_{1n}\right)\right| \leqslant \lim_{n\to\infty} ||P_{1n}'P_{2n} + P_{2n}'P_{1n} + P_{2n}'P_{2n}||_2 = 0. \tag{B.25}$$

As a result, if $\limsup_{n\to\infty} ||\rho_{m(n)} - \rho_0||_2 > 0$, by Assumption 3.5(iii),

$$\liminf_{n\to\infty} vareig\left[(P_{1n}+P_{2n})'(P_{1n}+P_{2n})\right] > 0.$$

Therefore, by the compactness of $\widetilde{\Theta}_{\sigma^2}$ and Lemma A.1,

$$\limsup_{n\to\infty} f\left(\frac{\sigma_0^2}{\sigma_{m(n)}^2}(P_{1n}+P_{2n})'(P_{1n}+P_{2n})\right) > 0, \tag{B.26}$$

contradicting equations (2.17) and (2.20). So, $\rho_{m(n)} \to \rho_0$. With this conclusion, again by equations (2.17) and (2.20), $\sigma_{m(n)}^2 \to \sigma_0^2$. That is to say, $\theta_{m(n)} \to \theta_0$, contradicting $\{\theta_{m(n)}\} \subseteq \widetilde{\Theta} - B(\theta_0, \gamma)$.

(iv) We will show that $\sup_{\theta\in\widetilde{\Theta}} \frac{1}{n} |\ln L_n(\theta) - E\ln L_n(\theta)| = o_p(1)$ in this step. Then condition (ii) in Lemma B.3 holds.

By Thm. 1 in Andrews (1992), it suffices to verify the pointwise convergence in probability and the stochastic equicontinuity of $\frac{1}{n}[\ln L_n(\theta) - E\ln L_n(\theta)]$ over $\widetilde{\Theta}$. In the following, we will verify these two conditions, respectively.

Notice that

$$l_{i,n}(\theta) - El_{i,n}(\theta) = \frac{1}{2\sigma^2}\left[\left\{R_{i\cdot,n}(\rho)[S_n(\lambda)Y_n - X_n\beta]\right\}^2 - E\left\{R_{i\cdot,n}(\rho)[S_n(\lambda)Y_n - X_n\beta]\right\}^2\right]$$

and

$$\left\{R_{i\cdot,n}(\rho)[S_n(\lambda)Y_n - X_n\beta]\right\}^2 = \left\{R_{i\cdot,n}(\rho)\left[S_n(\lambda)S_n^{-1}(X_n\beta_0 + V_n) - X_n\beta\right]\right\}^2$$
$$= V_n'\left[R_{i\cdot,n}(\rho)S_n(\lambda)S_n^{-1}\right]'\left[R_{i\cdot,n}(\rho)S_n(\lambda)S_n^{-1}\right]V_n + \left[R_{i\cdot,n}(\rho)S_n(\lambda)S_n^{-1}X_n\beta_0 - X_n\beta\right]'V_n +$$
$$\left[R_{i\cdot,n}(\rho)S_n(\lambda)S_n^{-1}X_n\beta_0 - X_n\beta\right]'\left[R_{i\cdot,n}(\rho)S_n(\lambda)S_n^{-1}X_n\beta_0 - X_n\beta\right].$$

For each $\theta \in \widetilde{\Theta}$, since $\left[R_{i\cdot,n}(\rho)S_n(\lambda)S_n^{-1}\right]'\left[R_{i\cdot,n}(\rho)S_n(\lambda)S_n^{-1}\right]$ is uniformly bounded in row sums, by Lemma A.12 in the supplement of Lee (2004), $\frac{1}{n}|\ln L_n(\theta) - E\ln L_n(\theta)| = o_p(1)$.

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

Direct calculations show that $E\sup_{\theta\in\widetilde{\Theta}}\left|\frac{1}{n}\frac{\partial}{\partial\theta_k}[\ln L_n(\theta)-E\ln L_n(\theta)]\right|=O(1)$. Therefore, by Lem. 1 in Andrews (1992), $\frac{1}{n}[\ln L_n(\theta)-E\ln L_n(\theta)]$ is stochastically equicontinuous on $\widetilde{\Theta}$.

(2) Next, we verify condition (3) of Lemma B.2 in three steps using Lemma B.4.

(i) First, we will show that for each $\theta\in\Theta$, the second-order derivatives of $l_{i,n}(\theta)$ are uniformly L_2-NED and uniformly $L_{2+\Delta/2}$ bounded. Similarly, for each $\theta\in\Theta$, the second-order derivatives of $El_{i,n}(\theta)$ are uniformly bounded.For each ρ, by Lemma 1, $\{R_{i\cdot,n}(\rho)W_{kn}Y_n\}$ is uniformly $L_{4+\Delta}$ bounded. As a consequence, for each θ, by the Cauchy–Schwarz inequality, $\{\frac{\partial^2 l_{i,n}(\theta)}{\partial\lambda_j\partial\lambda_k}\}$ is uniformly $L_{2+\Delta/2}$ bounded. So are the remaining terms in the Hessian matrix of $l_{i,n}(\theta)$ by similar arguments.

(ii) Second, we will show that $\Omega_n^{-1/2}\frac{1}{\sqrt{n}}\frac{\partial\ln L_n(\theta_0)}{\partial\theta}\xrightarrow{d}N(0,I)$. The expressions of $\frac{\partial\ln L_n(\theta_0)}{\partial\theta}$ are $\frac{\partial\ln L_n(\theta)}{\partial\lambda_j}=-tr(S_nW_{jn})+\frac{1}{\sigma^2}V_n'R_nW_{jn}S_n^{-1}(X_n\beta^0+R_n^{-1}V)$, $\frac{\partial\ln L_n(\theta)}{\partial\beta}=\frac{1}{\sigma^2}(R_nX_n)'V_n$, $\frac{\partial\ln L_n(\theta)}{\partial\sigma^2}=-\frac{1}{2\sigma^2}+\frac{1}{2\sigma^4}V_n'V_n$, $\frac{\partial\ln L_n(\theta)}{\partial\rho_k}=-tr\left[R_n(\rho)^{-1}M_{kn}\right]+\frac{1}{\sigma^2}V_n'M_{kn}R_n^{-1}V_n$.
Consider $\frac{\partial\ln L_n(\theta)}{\partial\lambda_j}$ first. Notice that

$$\sup_{i,n}\left|(R_nW_{jn}S_n^{-1}X_n\beta^0)_i\right|=\sup_{i,n}\left|R_{i\cdot,n}W_{jn}S_n^{-1}X_n\beta^0\right|=\sup_{i,n}\left\|R_{i\cdot,n}W_{jn}S_n^{-1}X_n\beta^0\right\|_\infty$$
$$\leqslant\sup_{i,n}\left(\left\|R_{i\cdot,n}\right\|_\infty\left\|W_{jn}\right\|_\infty\left\|S_n^{-1}\right\|_\infty\left\|X_n\beta^0\right\|_\infty\right)<\infty. \tag{B.27}$$

Similarly, $\sup_n\left\|R_nW_{jn}S_n^{-1}R_n^{-1}\right\|_\infty<\infty$, $\sup_n\|R_nX_n\|_\infty<\infty$, and $\sup_n\left\|M_{kn}R_n^{-1}\right\|_\infty<\infty$. Then by Assumption 3.6, Lem. A.13 in the supplement of Lee (2004), and the Cramér–Wold theorem, $\Omega_n^{-1/2}\frac{1}{\sqrt{n}}\frac{\partial\ln L_n(\theta_0)}{\partial\theta}\xrightarrow{d}N(0,I)$.

(iii) Finally, we aim to verify the uniform convergence in probability condition (iii) in Lemma B.4. According to Thm. 1 in Andrews (1992), it suffices to verify the pointwise convergence in probability and the stochastic equicontinuity of $\frac{1}{n}\left[\frac{\partial^2\ln L_n(\theta)}{\partial\theta_j\partial\theta_k}-E\frac{\partial^2\ln L_n(\theta)}{\partial\theta_j\partial\theta_k}\right]$ in a neighborhood of θ_0.

Since every term in the Hessian matrix of $\ln L_n(\theta)$ is of a linear-quadratic form of V_n, by Lem. A.12 in the supplement of Lee (2004), the pointwise convergence in probability holds: for each $\theta\in\Theta$, $\frac{1}{n}\left[\frac{\partial^2\ln L_n(\theta)}{\partial\theta_j\partial\theta_k}-E\frac{\partial^2\ln L_n(\theta)}{\partial\theta_j\partial\theta_k}\right]=o_p(1)$.

Because $\theta\in\Theta^0$, there is a $\delta>0$ such that $\sup_{\lambda\in[\lambda_0-\delta,\lambda_0+\delta],n}\left\|S_n(\lambda)^{-1}\right\|_\infty<\infty$ and

$$\sup_{\rho\in[\rho_0-\delta,\rho_0+\delta],n}\left\|R_n(\rho)^{-1}\right\|_\infty<\infty.$$

To establish the stochastic equicontinuity of $\frac{1}{n}\left[\frac{\partial^2\ln L_n(\theta)}{\partial\theta_j\partial\theta_k}-E\frac{\partial^2\ln L_n(\theta)}{\partial\theta_j\partial\theta_k}\right]$ on $\overline{B(\theta_0,\delta)}\equiv\{\theta:||\theta-\theta_0||\leqslant\delta\}\subset\Theta$, by Lem. 1 in Andrews (1992), it is sufficient to show that for all j, k,

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press

and l,

$$E \sup_{\theta:||\theta-\theta_0||\leqslant\delta} \left| \frac{1}{n} \left[\frac{\partial^3 \ln L_n(\theta)}{\partial\theta_j\partial\theta_k\partial\theta_l} - E\frac{\partial^2 \ln L_n(\theta)}{\partial\theta_j\partial\theta_k\partial\theta_l} \right] \right| = O(1). \tag{B.28}$$

For the nonstochastic terms , for example, $\frac{1}{n}\frac{\partial^3 \ln L_n(\theta)}{\partial\lambda_k\partial\lambda_j\partial\lambda_r}$ and $\frac{1}{n}\frac{\partial^3 \ln L_n(\theta)}{\partial\lambda_k\partial\lambda_j\partial\beta}$, equation (2.28) holds trivially. $\frac{1}{n}\frac{\partial^3 \ln L_n(\theta)}{\partial\lambda_k\partial\lambda_j\partial\sigma^2}$: Since $\sup_{i,n,\,\rho\in[\rho_0-\delta,\,\rho_0+\delta]} ||R_{i,n}(\rho)W_{jn}||_\infty < \infty$ and $\sup_{i,n} ||y_{i,n}||_{L^{4+\Delta}} < \infty$, by Minkowski's inequality, $\sup_{i,n,\,\rho\in[\rho_0-\delta,\,\rho_0+\delta]} ||R_{i,n}(\rho)W_{jn}Y_n||_{L^{4+\Delta}} < \infty$. As a result, by the Cauchy–Schwarz inequality, $\sup_{i,n,\,\theta\in\overline{B(\theta_0,\,\delta)}} \left\| \frac{1}{n}\left[\frac{\partial^3 l_{i,n}(\theta)}{\partial\theta_j\partial\theta_k\partial\theta_l} - E\frac{\partial^2 l_{i,n}(\theta)}{\partial\theta_j\partial\theta_k\partial\theta_l} \right] \right\|_{L^{2+\Delta/2}} < \infty$, and equation (2.28) holds for $\frac{1}{n}\frac{\partial^3 \ln L_n(\theta)}{\partial\lambda_k\partial\lambda_j\partial\sigma^2}$.

For the remaining terms, equation (2.28) holds by similar arguments. □

Proof of Theorem 2: We prove this theorem by verifying the conditions of Lemma B.2.

(1) The uniform convergence in probability condition $\sup_{\theta\in\Theta} \frac{1}{n} |\ln L_n(\theta) - E\ln L_n(\theta)| = o_p(1)$ has been verified in the proof of Thm. 1 in Xu and Lee (2015). With Assumption 3.13, by Lemma B.3, condition (3) in Lemma B.2 holds.

(2) Conditions (ii) and (iii) of Lemma B.4 are verified in the proof of Thm. 2 in Xu and Lee (2015). Thus, by Lemma B.4, condition (4) in Lemma B.2 is satisfied.

Hence, by Lemma B.2, we have the conclusion of the theorem. □

https://doi.org/10.1017/S0266466620000286 Published online by Cambridge University Press